DOCUMENT RESUME

ED 071 761 PS 006 305

AUTHOR Montgomery, Mary Eleanor
TITLE The Interaction of Three Levels of Aptitude Determined by a Teach-Test Procedure with Two Treatments Related to Area. Report from the Project on Development of Instructional Programs: Analysis of Mathematics Instruction.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-5-0216
PUB DATE Sep 72
CONTRACT OEC-5-10-154
NOTE 145p.; Technical Report No. 235

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS Aptitude Tests; Behavioral Objectives; Concept Teaching; *Curriculum Development; *Evaluation Techniques; Interaction; Primary Grades; *Problem Solving; *Relevance (Education); Research; Test Results; *Tests

ABSTRACT

The general problem addressed in this thesis concerns formative evaluation relevant to curriculum development. The research strategy was that of an aptitude-treatment interaction (ATI) study. Aptitude was defined in terms of the individual's ability to learn specific concepts associated with a unit of length measurement. The treatments were designed to differ only in their emphasis on a unit of area measurement. The specific question asked was: In what manner does the ability of children to learn concepts associated with a unit of length affect the extent to which they attain concepts associated with area and a unit of area for each of the two given treatments? In order to determine this ability, 90 second and third graders were subjected to a teach-test procedure. This procedure consisted of a pretest, a brief instructional treatment and a posttest, all of which tested or taught about a unit of length. The results of the two tests were used to determine the aptitude levels. No significant interactions were found between the aptitudes and treatments on any of the measures. There were significant main effects due to aptitude and to treatment for achievement and retention measures. Other findings relevant to curriculum development reported in this study are: (1) It is feasible to teach these area concepts to second and third graders, and (2) Second and third graders are capable of handling conflicting situations involving units of area.
(Author/CK)

FILMED FROM BEST AVAILABLE COPY

BR-5-0216

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 071761

Technical Report No. 235

THE INTERACTION OF THREE LEVELS OF APTITUDE DETERMINED BY A TEACH-TEST PROCEDURE WITH TWO TREATMENTS RELATED TO AREA

Report from the Project on Development of Instructional Programs: Analysis of Mathematics Instruction

by Mary Eleanor Montgomery

Thomas A. Romberg and John G. Harvey
Principal Investigators

PS 006305

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

September 1972

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

STATEMENT OF FOCUS

Individually Guided Education (IGE) is a new comprehensive system of elementary education. The following components of the IGE system are in varying stages of development and implementation: a new organization for instruction and related administrative arrangements; a model of instructional programing for the individual student; and curriculum components in prereading, reading, mathematics, motivation, and environmental education. The development of other curriculum components, of a system for managing instruction by computer, and of instructional strategies is needed to complete the system. Continuing programmatic research is required to provide a sound knowledge base for the components under development and for improved second generation components. Finally, systematic implementation is essential so that the products will function properly in the IGE schools.

The Center plans and carries out the research, development, and implementation components of its IGE program in this sequence: (1) identify the needs and delimit the component problem area; (2) assess the possible constraints--financial resources and availability of staff; (3) formulate general plans and specific procedures for solving the problems; (4) secure and allocate human and material resources to carry out the plans; (5) provide for effective communication among personnel and efficient management of activities and resources; and (6) evaluate the effectiveness of each activity and its contribution to the total program and correct any difficulties through feedback mechanisms and appropriate management techniques.

A self-renewing system of elementary education is projected in each participating elementary school, i.e., one which is less dependent on external sources for direction and is more responsive to the needs of the children attending each particular school. In the IGE schools, Center-developed and other curriculum products compatible with the Center's instructional programing model will lead to higher morale and job satisfaction among educational personnel. Each developmental product makes its unique contribution to IGE as it is implemented in the schools. The various research components add to the knowledge of Center practitioners, developers, and theorists.

ACKNOWLEDGEMENTS

There are many people who deserve expressed appreciation for their contribution to this thesis and to my education. It is impossible to recognize all of them or to recognize properly any of them.

Dr. Thomas A. Romberg, my major professor, has provided the direction and encouragement necessary not only for this thesis, but also for many other aspects of my graduate work: to him, my unqualified thanks. I would also like to thank Professor John G. Harvey for his assistance in editing as well as for serving on my committee and Professors J. Fred Weaver, Frank H. Farley, and Simon Hellerstein for serving on my committee.

No acknowledgement would be complete without thanking those involved in the actual study. Miss Marcia Dana gave willingly of her time and talents as she prepared to teach and as she taught the experimental unit. The principal, Mr. Kness, the second and third grade teachers, and the students who participated made Hartland Elementary School an ideal place to conduct a study.

The Wisconsin Research and Development Staff, especially those associated with Developing Mathematical Processes, all are due thanks.

Finally, to my mother, whose creative talents and faith in her children have been an inspiration, and to my father, whose ideals of education and positive outlook have been a steady influence: accept this as a token of my appreciation.

TABLE OF CONTENTS

PAGE

ACKNOWLEDGEMENTS....iv

LIST OF TABLES....viii

LIST OF FIGURES....ix

ABSTRACT....xi

CHAPTER

I. INTRODUCTION TO THE THESIS....1

Introduction....1
Crucial Terms....1
From the General to the Specific Problem....5
Brief Description of the Experiment....7
Outline of Remainder of the Thesis....8

II. THE PROBLEM....10

Introduction....10
General Problem....10
Background....11
DMP's approach to mathematics....12
The process of measuring....13
The role of the unit....16
Specific Problem....17
Research Strategy....19
Treatments....20
Aptitudes....22
Specific Questions....27

III. RELEVANT AND RELATED RESEARCH....28

Introduction....28
Relevant Measurement Literature....29
The role of the unit....29
The relation between measuring
length and measuring area....33
The measuring of area....35

CHAPTER PAGE

Aptitude-Treatment Interaction Literature.....37
Teach-test Literature.....41
Summary.....44

IV. DEVELOPMENT OF THE STUDY.....46

Introduction.....46
Development of the Problem.....48
Formulation of the original problem.....48
Analysis for the Treatments A and B.....48
Pilot studies of Treatments A and B.....48
Reformulation of the problem.....50
Development of the Teach-test Procedure.....51
Analysis for teach-test procedure.....51
Pilot Study #1 of the teach-test procedure.....52
Reanalysis #1 for the teach-test procedure.....53
Pilot Study #2 of the teach-test procedure.....54
Reanalysis #2 for the teach-test procedure.....54
The final plan for the teach-test procedure.....55
Development of the Treatments.....57
Analysis for the treatments.....57
The final plan for the treatments.....60
The Research Strategy.....67
Research design.....67
Population.....69
Hypotheses.....69
Proposed statistical analysis.....71

V. EXECUTION OF THE STUDY.....74

Introduction.....74
Population.....74
Prestudy Arrangement.....75
Selection of teacher.....75
Initial contact with the school.....75
Meeting with the teachers involved.....76
Summary of the Teach-test Procedure.....77
Results of $O_1 - O_2$.....78
Determination of Aptitude Levels.....80
Treatments' Population.....86
Summary of Treatments.....87
Results of $O_3 - O_5$.....89

VI. DATA ANALYSIS.....92

Introduction.....92
Achievement Measure.....92

CHAPTER PAGE

Transfer Measure....96
Retention Measure....99
Retention Ratios....102
Summary....103

VII. CONCLUSION TO THE THESIS....105

Introduction....105
Summary....105
Limitations....107
Conclusions....108
Implications for Curriculum Development....116
Recommendations for Future Research....124
Concluding Remarks....126

REFERENCES....129

LIST OF TABLES

TABLE		PAGE
4.1	Behavioral Objectives for Treatments U and N	58
4.2	Instructional Objectives for Treatments U and N	61-62-63
5.1	Descriptive Statistics for O_1 and O_2	79
5.2	Descriptive Statistics for Levels I and II	86
5.3	Descriptive Statistics for O_3 - O_5	90
6.1	Descriptive Statistics for O_3: Aptitude, Treatment and Aptitude by Treatment Groups	93
6.2	Analysis of Variance for O_3 (Achievement)	95
6.3	Descriptive Statistics for O_4: Aptitude, Treatment and Aptitude by Treatment Groups	96
6.4	Analysis of Variance for O_4 (Transfer)	98
6.5	Descriptive Statistics for O_5: Aptitude, Treatment and Aptitude by Treatment Groups	99
6.6	Analysis of Variance for O_5 (Retention)	101
6.7	Individual Results on Achievement (O_3) and Retention (O_5) and Retention Ratios	103
7.1	Group Means on Achievement (O_3) and Retention (O_5) and Retention Ratios	113

LIST OF FIGURES

FIGURE PAGE

2.1 Flow Chart for the Attribute of Area.....23

2.2 Flow Chart for the Attribute of Length.....24

4.1 Development of Experiment.....47

4.2 Research Design.....68

5.1 Physical and Temporal Arrangements for Treatments.....77

5.2 Criteria for Aptitude Levels.....82

5.3 Possible Combinations of Mastery Levels on O_1 and O_2 and Classifications of These Combinations.....84

5.4 Description of Treatments' Sample According to Level, Age, Sex and Grade.....87

6.1 Aptitude Effect on Achievement.....94

6.2 Treatment Effect on Achievement.....94

6.3 Aptitude Effect on Transfer.....97

6.4 Treatment Effect on Transfer.....97

6.5 Aptitude Effect on Retention.....100

6.6 Treatment Effect on Retention.....100

7.1 Classifications for Aptitude Levels Based on Subscale 2 of O_1 and O_2.....110

7.2 Mean Scores on Achievement According to Two Classification Schemes.....111

7.3 Percent of Correct Responses on Items 4-7 of Observation O_4.....116

FIGURE PAGE

7.4 Number of Each Type of Comparison Problems Which Involved Two Covered Regions on Observation O_3.....119

7.5 Percent of Correct Responses to Comparison Problems Which Involved Two Covered Regions on Observation O_3.....120

The project presented or reported herein was performed pursuant to a Grant from the United States Office of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the United States Office of Education, and no official endorsement by the United States Office of Education should be inferred.

ABSTRACT

The general problem addressed in this thesis concerns formative evaluation relevant to curriculum development. More specifically, the process of measuring as it relates to primary students was the portion of mathematics curriculum considered. This was narrowed to examining the interaction of two treatments on measuring area with various levels of aptitude. The research strategy was that of an aptitude-treatment interaction (ATI) study.

Most of the past educational ATI studies prescribe treatments which differ on form or mode and capitalize on the learners' best general abilities. This study approached the ATI question in a different manner. General capabilities were not used as measure of aptitude and the treatments did not differ in form or mode. Aptitude was defined in terms of the individual's ability to learn specific concepts associated with a unit of length measurement. The treatments were designed to diff only in their emphasis on a unit of area measurement. The specific question asked was: In what manner does the ability of children to learn concepts associated with a unit of length affect the extent to which they attain concepts associated with area and a unit of area for each of the two given treatments?

In order to determine this ability 90 second and third graders were subjected to a teach-test procedure. This procedure consisted of a pretest, a brief instructional treatment and a posttest all of which tested or taught about a unit of length. The results of the two tests

were used to determine the aptitude levels. Although three levels of aptitude were expected only two subjects met the criteria for one of the levels. They were dropped from the remainder of the study along with those students who did not fit the definition of either of the other two levels. The students in each of the other two levels (32 in Level I and 27 in Level II) were randomly assigned to the two treatments.

Both treatments had the same behavioral objectives, the same teacher, the same duration (9 days) and the same mode of instruction. They differed in their treatment of the unit of measure for area. After the treatments three measures, achievement, transfer and retention, were taken. These measures were used to test hypotheses about the interaction of aptitude with treatments and about the main effects of aptitude and of treatment.

No significant interactions were found between the aptitudes and treatments on any of the measures. There were significant main effects due to aptitude and to treatment for achievement and retention measures. In interpreting these results one must take into account that one level of aptitude was not found in the given population. It had been hypothesized that much of the interaction would have been due to this level.

Other findings relevant to curriculum development reported in this study are: 1) It is feasible to teach these area concepts to second and third graders, 2) Second and third graders are capable of handling conflicting situations involving units of area.

Chapter I

INTRODUCTION TO THE THESIS

Introduction

The main purpose of this study was to examine the interaction of two treatments on measuring area with various levels of aptitude. This first sentence contains many words (i.e., area, measuring and aptitude) which may be interpreted in a variety of ways. To minimize misunderstandings or misinterpretations throughout this thesis the first chapter begins by defining or explaining crucial terms.

The remainder of this chapter gives an overview of the study and of the thesis. Both the general problem considered and the specific problem investigated are identified and a brief description of the experiment is included. To complete this introductory chapter the remaining chapters are outlined.

Crucial Terms

Aptitude: "Any characteristic of the individual that increases (or impairs) his probability of success in a given (educational) treatment." (Cronbach and Snow, 1969, p. 7) The characteristic examined in this study was the child's ability to learn certain mathematical concepts about a unit of length. This ability was determined through a procedure known as a teach-test procedure.

Aptitude Levels. Three levels of aptitude were used to classify subjects as follows:

Level I — Subjects who had not attained the specified behaviors by the end of the teach-test procedure and evidenced no change in performance from pretest to posttest.

Level II — Subjects who attained the specified behaviors only after the teach-test procedure and evidenced change in performance from pretest to posttest.

Level III — Subjects who had already attained the specified behaviors before and maintained them throughout the teach-test procedure.

Aptitude-Treatment Interaction (ATI): The interaction between learning abilities (aptitudes) and instructional treatments. Studies which investigate this interaction are known in the literature as ATI studies.

Attributes: Characteristics or properties of objects or sets. This study is concerned only with the attributes of length and area and only these as they are described here.

Length: An undefined attribute; perceptually the dominant attribute of long, thin objects and mathematically represented by a line segment.

Area: An undefined attribute; perceptually a dominant attribute of flat objects and mathematically represented by a planar region.

Curriculum. A broad view of curriculum is taken in this thesis. Mathematics curriculum is composed of four components: mathematics program, learner, teacher, and instruction (Romberg and DeVault, 1967) and is defined by specifying the components and the interrelations among them.

Developing Mathematical Processes (DMP). A K-6 elementary mathematics program being developed by the Analysis of Mathematics Instruction (AMI) Project at the Wisconsin Research and Development Center for Cognitive Learning. The program is based on a measurement approach to mathematics and mathematical processes are emphasized. The instructional treatments for this study were developed to reflect this approach to mathematics (Romberg, Fletcher and Scott, 1968).

Measure-Measurement. In this thesis a distinction is made between measure and measurement. A measurement is a symbolic representation of an attribute which includes both the unit and the number assigned by the process of measuring. A measure is a symbolic representation of an attribute consisting only of the number.

Processes. In this study only the following processes as defined in DMP were emphasized.

Comparing: The process of deciding whether two objects are alike with respect to a stated attribute.

Ordering: The process of deciding the direction of the difference between two objects with respect to a stated attribute.

Representing: The process of denoting or expressing in a different medium an attribute of an object or a set. If the medium is physical, a physical representation is made. Similarly, if the medium is a picture or a symbol then a pictorial or symbolic representation is made.

Measuring: The process of assigning a numeral as a symbolic representation of an attribute of an object. These steps are followed:

1) A set of objects is recognized as possessing the attribute,
2) the procedures for comparing, ordering, and combining the objects are established, and
3) a unit is specified, thus assigning to each object of the domain a non-negative number.

Teach-test Procedure (T-T): A particular procedure which simulates the actual classroom environment in order to attain a measure of the student's ability to learn. The procedure consists of a pretest, a brief instructional unit, and a posttest. In this study this procedure was used to determine the aptitude levels.

Treatments. There were two treatments designed for the study. No attempt is made here to describe fully the treatments; the intent of this statement is only to indicate that Treatment U emphasizes

the unit and Treatment N does not. For a complete description of the treatments see Chapter V and Appendix C.

From the General to the Specific Problem

"To generate knowledge about mathematics instruction and to incorporate it into a validated instructional program" (Romberg and Harvey, 1969, p. 1) is the stated purpose of the project, Analysis of Mathematics Instruction, which is developing the program Developing Mathematical Processes (DMP). This curriculum program approaches elementary number concepts through measurement. This is not a common approach and little research exists which is relevant to some of the problems posed by this approach. This study is part of a series of studies (Scott, 1969; Gilbert, 1969; Weinstein, 1970; Carpenter, 1971; and Carpenter, in press) which are designed to generate knowledge about measurement instruction so that it may be incorporated into DMP.

PS 010 605

At the time this study was being planned and executed, instructional topics on area were being developed and questions about the child's understanding of the role of a unit in the process of measurement were being raised. The research on area and on the role of the unit is even more scarce than research on many other measurement questions. Thus, the investigator decided to examine the role of the unit as it applied to measuring area.

Furthermore, DMP is the mathematics component of Individually

Guided Education (IGE), the focus of the Wisconsin Research and Development Center for Cognitive Learning (Klausmeier, Quilling, and Sorenson, 1971). In developing an individually guided mathematics program based on research it seems natural to respond to Cronbach and Snow's (1969) pleas for research which formulates more precisely the ways in which programs can be varied so as to fit learners' characteristics. Therefore, the method of research adopted to investigate the role of a unit in measuring area was an aptitude-treatment interaction study.

Most of the past educational ATI studies fit what Salamon (1971) has described as the preferential model. This model prescribes treatments differing on form or mode and capitalizes on the learners' best general capabilities. However, few of these studies have been successful in producing the desired interaction (Bracht, 1969).

This study approached the ATI question in a different manner. General capabilities were not used as measures of aptitude and the treatments did not differ on form or mode. Aptitude was defined in terms of the individual's ability to learn specific concepts associated with a unit of length measurement. The treatments were designed to differ on their emphasis on the unit of area measurement. Thus, the specific question asked was: In what manner does the ability of children to learn concepts associated with a unit of length affect the extent to which they attain concepts associated with area and a unit of area for each of the two given treatments?

Brief Description of the Experiment

The basic research design of an aptitude-treatment interaction experiment is one in which the identified aptitude is measured, the subjects are randomly assigned to a treatment, and the outcome is measured (Cronbach and Snow, 1969, p. 21).

Therefore, the first step after identifying the aptitude is to measure it. Since aptitude was defined as the child's ability to attain certain behaviors concerning a unit of length, the following procedure (the teach-test procedure) was used to measure aptitude. On the stated objectives concerning length 110 second and third graders in one school were given a pretest, were instructed for two days, and were given a posttest. From the results of the tests, subjects were classified into two levels: Level I, those students who had not attained the specified behaviors prior to or after instruction and who evidenced no change in performance; Level II, those students who had not attained the specified behaviors prior to instruction, but had attained them after the instruction and who had evidenced change in performance. Because of previous experimentation a third level of aptitude, those students who had attained the specified behaviors prior to instruction, was expected. Only two children in the experiment's population fitted this description and, therefore, were dropped from the study. Those students who were absent during the testing or instruction or who did not fit in either Level I or Level II were also dropped. (A complete description of the criteria used to determine the levels is found in Chapter V.)

The remaining subjects in Level I and in Level II were assigned randomly to two treatments. The two treatments were designed to interact with the aptitude. Since the aptitude reflected the ability to learn concepts about a unit of length, one treatment (Treatment U) was developed to emphasize similar concepts about a unit of area and the other (Treatment N) was developed to de-emphasize such concepts. Both treatments were activity oriented, were taught by the same teacher, were monitored by the experimenter and consisted of nine forty-minute sessions.

Upon completion of the instructional treatments three measures, achievement, transfer and retention, were taken. The results, analysis and implications are discussed in full in the last three chapters.

Outline of the Remainder of the Thesis

Chapter II demonstrates the significance of the general problem and draws a relation of the specific problem to the general problem. The discussion of the specific problem includes an analysis of it, the rationale, the proposed questions, and its significance. Chapter III is a summary of research directly relevant or indirectly related to the study.

Chapter IV contains the considerations made in designing and the plans made for conducting the study. The pilot studies and their effects on the plans for the final study are reported. The design,

the population, the teach-test procedure, the treatments, the instruments, the hypotheses and the analysis are presented as they were proposed for the study.

In Chapter V the actual execution of the experiment is described. The population is characterized, the summary of the teach-test procedure is given, the results of the teach-test observations are reported, the aptitude levels are defined, the daily description of the treatments is included, and the results of the treatment observations are reported.

Chapter VI is a report of the statistical analysis. Chapter VII, the concluding chapter, includes interpretations of the analysis, a summary of the study and its implications, and the projections for further study.

Chapter II

THE PROBLEM

Introduction

The specific question asked in this study was: How does children's ability (aptitude) to learn concepts about a unit of length interact with alternative treatments on measuring area? In this chapter the derivation of this question from the general problem is reported. First the general problem and its significance is identified. After the necessary background for understanding the specific problem is given, the specific problem is identified. Next a description of the research strategy used to investigate the specific problem and its influence on the final question is given. The treatments are briefly described and the method for determining the aptitude is explicated so that the specific question may be interpreted accordingly.

General Problem

"Evaluation, used to improve the course while it is still fluid, contributes more to improvement of education than evaluation used to appraise a product once it is on the market." (Cronbach, 1969, p. 364) Many similar quotes by those who promote what Scriven (1967) has called formative evaluation may be found, but few reports of such studies are

available in published literature. This partly is due to one function of formative studies--the changes are made in the product according to the findings and these constitute the report of the study.

However, with the increase of federal and private corporations' monies for research and development, more encouragement is being given formative studies of curriculum projects (Kirst and Walker, 1971) and the reports thereof. The products being developed at Research and Development Centers or at Regional Laboratories fall into this category. One such product is Developing Mathematical Processes (DMP), a mathematics program for grades K-6, being developed by the mathematics project (Analysis of Mathematics Instruction Project) at the Wisconsin Research and Development Center for Cognitive Learning.

The underlying purpose of the Analysis of Mathematics Instruction Project, "to generate knowledge about mathematics instruction and to incorporate it into a validated mathematics program," (Romberg and Harvey, 1969) establishes the need for such research in connection with this project. The investigator, working for the project, assumed this need and the significance thereof, and proceeded to look at a portion of this general problem--the portion concerning instruction in the process of measuring.

Background

To understand the specific problem it is necessary to have some background of DMP's approach to mathematics, the process of measuring,

and the role of the unit in measuring.

DMP's approach to mathematics. DMP approaches elementary mathematics through measurement (Romberg, Fletcher, and Scott, 1968). This is not a common approach; most contemporary elementary mathematics programs use a set theoretical approach. That is, most programs begin with counting sets of objects to attain mastery of number concepts, but DMP begins with the processes underlying measuring and eventually a measure (number) is assigned to a given attribute of an object or of sets.

These processes underlying measuring--comparing, ordering, and representing--are three of the processes emphasized in DMP. The program begins in kindergarten, after a topic in which objects are described and classified on attributes familiar to children, by focusing on the attribute of length--an attribute perceptually salient to children of this age. First, the children compare and order lengths directly by placing the objects side by side. When these processes are mastered in this context, the children are presented the problem of comparing two objects which cannot be placed side by side. The process of physically representing a length is introduced to help them solve this problem. In refining this process, the procedure of laying units end to end in order to represent a length is learned. The children proceed to pictorially representing lengths and, finally, to symbolically representing lengths. This last step is taken late in kindergarten or early in the first grade after the same processes

have been considered with the attribute of numerousness. Throughout the program these same processes, along with others, are reconsidered with many other attributes--weight, capacity, duration, area, etc.

Symbolic stimuli which may be interpreted as representing attributes or processes are not given until after the children have been exposed to related physical or pictorial stimuli. Thus, at the early levels DMP depends heavily on the child's ability to model a symbolic statement physically or pictorially and not upon manipulating symbols.

The measurement approach complements DMP's other pedagogical strategies--use of physical referents, problem-solving, functional transition, and spiraling techniques (Romberg, Fletcher, and Scott, 1968). It provides a wide variety of physical referents by considering attributes other than just numerousness. It allows the children to solve problems with perceptually meaningful materials; problems similar to those they will later be presented symbolically. Many of the functional transitions to a new process or to a new skill can be made through measurement problems which the child can solve at an early age. The introduction of new attributes permits spiraling through the processes at a pace suitable to the child's development.

This description of DMP is not intended to be exhaustive; it only includes those points relevant to the specific problem. For a complete description the following papers may be referenced (Romberg and Harvey, 1969; Harvey, Romberg, and Fletcher, 1969).

The process of measuring. The process of measuring may be con-

sidered as one of mapping empirical properties or relations into a formal model (Stevens, 1959, p. 20). This definition requires a set A of elements which possesses the properties and relations under consideration, a set B which possesses a formal structure, and a procedure for associating each element of A with an element of B in such a way that the essential properties and relations of set A are preserved (Blakers, 1967). While a more precise definition is necessary for the study of all measurable functions, this definition is sufficient for most common measure functions. Since the measure functions in this study are related to length and area, the remainder of this discussion is relevant to such common functions.

The first step in the measuring process is identification of a domain of elements which possesses the attribute under question. (For example, the domain may be the set of all objects which possess the attribute of length.) Through empirical procedures a recognizable structure is imposed upon this domain. An equivalence relation is established through a procedure for comparing the elements. (For the attribute of length this amounts to deciding whether or not the objects are the same length.) This equivalence relation partitions the domain into equivalence classes. Further, a strict total order relation is established through the procedure of ordering elements from the equivalence classes. (For the attribute of length two equivalence classes may be ordered by ordering representative objects from each class, that is, by deciding which object is longer.) Lastly, a binary operation is defined which is associative and commutative.

(In the case of length, this operation may be described as "combining" or joining two objects end to end.) Through these relations and this operation the recognizable structure of an ordered abelian semi-group is imposed upon the domain.

The next step in the process of measuring is specifying the formal model, that is, indicating a set which possesses a formal structure. For common measure functions this set is the non-negative real numbers. The only step remaining is the definition of a function which preserves the essential characteristics of the domain and assigns each element of the domain to a non-negative real number. This assignment is completely defined by specifying a unit, that is, by identifying an equivalence class whose image is 1. Thus, an isomorphism is set up between the domain and the set of non-negative real numbers which preserves the operation and order relation imposed on the domain.

Therefore, in the process of measuring length, area, or any other common measurable attribute the same steps are followed:

1) A set of elements (the domain) is recognized as possessing the attribute,
2) the procedures for comparing, ordering, and combining the elements of the domain are established, and
3) a function is defined which preserves the essential characteristics of domain and assigns to each element of the domain a non-negative real number.

This study is concerned with all three steps, although the function is always defined simplistically as the counting function. That is, each element of the domain is mapped onto a non-negative integer by counting the number of units combined so as to be equivalent (to the nearest unit) to that element.

The role of the unit. As stated in the previous section, the assignment of each element to a non-negative real number is defined by specifying an equivalence class to be the unit. Technically, each time a different equivalence class is specified to be the unit, a different function is defined. For example, if the unit of length is specified to be two inches, then this measure function maps all lengths of two inches onto 1, all lengths of twelve inches onto 6, etc. But, if the unit of length is chosen to be four inches, then this function maps all lengths of four inches onto 1, all lengths of twenty-four inches onto 6, etc.

If in comparing two measures one fails to realize that different functions may have been used in assigning the measures, that is, different units have been specified, then one may reach an erroneous conclusion. For example, if the fact that two different units have been used in the previous examples is ignored, the conclusions may be reached that a length of twelve inches is equal to a length of twenty-four inches since they both correspond to 6 when mapped (by the first and second functions, respectively, as defined in the previous examples) onto the non-negative reals.

The Russians, Gal'perin and Georgiev (1969), addressed themselves to this problem. They proposed a program which began the study of numbers by focusing on the unit. Their experimentation found that the experimental group who were taught about units achieved superior results on the 14 measurement and conservation tasks administered.

Not many elementary series in this country approach this problem. Most of their problems are structured so that only one function is being used at a time. This procedure is not consistent with DMP's approach of letting the children generate their own numbers and problems, since then there is no guarantee that only one function will arise within a problem.

Specific Problem

In the process of developing a curriculum there are many decisions to be made. Some of these decisions are made subjectively, some are derived logically from basic assumptions, and some are made using empirical evidence. While the investigator was working on the development of DMP, the need to examine the attribute of area arose. Although the decision to include area after the attributes of length, numerousness, and weight and to treat area in the same manner as these attributes had been treated, no firm decisions had been made about the placement or the particulars of the activities in the total instructional program.

In late spring of 1971 the investigator worked with first and

second graders on activities concerning area. Both a manipulative approach and a more traditional rule-example approach were tried. At the end of the activities (a more complete description is given in Chapter IV), both groups could exhibit the specified behaviors. However, in individually testing the subjects, the investigator noted that only some children from both groups had difficulty on the transfer items which asked for a comparison of two regions not covered with a common unit. Some children focused only on the number of units, but others coordinated the size of the unit and the number. Likewise, when measuring a region covered with non-congruent units, some children responded only with the number of pieces but others coordinated the relationships between the pieces to respond correctly. If questioned more thoroughly about the unit, some children who initially responded incorrectly then focused on the unit and responded correctly.

These responses were not too surprising; other studies (Carpenter, 1971; Gal'perin and Georgiev, 1969; and Piaget, 1964) have reported similar responses. However, they did clearly point out that some children are ready to assign a number to area and others need more experience with the unit.

Working within the Individually Guided Education (IGE) framework of the Wisconsin Research and Development Center, the next natural questions were: (1) How to determine those individuals who needed and those who did not need more experience with the unit and (2) what

treatments would be necessary for each group.

Thus, the specific problem was raised: How does children's ability to learn specified concepts about a unit of measurement interact with alternative treatments, one of which emphasizes the unit and the other of which does not?

In order to answer this question the research strategy of an aptitude-treatment interaction study was invoked. The next section describes this strategy and its influence on this study.

Research Strategy

This specific problem is an instance of the general aptitude-treatment interaction question: "In what manner do the characteristics of learners affect the extent to which they attain the outcomes for each of the treatments that might be considered?" (Cronbach and Snow, 1969, p. 6)

Such a strategy calls for specifying aptitudes and designing treatments which interact with these aptitudes. That is, a treatment A is designed that is better for a learner with given characteristics and a treatment B is designed that is better for a learner with other characteristics. For a complete description of this strategy the summary by Cronbach and Snow (1969) is referenced.

Although many supporters of individual differences have prompted this strategy, few educational studies have produced the desired interaction. Bracht (1969) analyzed 90 ATI studies in terms of their

treatments, aptitudes (personological variables), and dependent variables in hopes of explaining the lack of disordinal interactions. However, only 5 of the possible 108 results of these studies showed disordinal interaction. Hence, only mild support was given to any hypotheses he made about the different types of variables. His work serves as a caution to those designing aptitude-treatment studies--the treatments must be carefully designed to take full advantage of the specified aptitudes.

Cahen (1969) urges those involved in ATI studies to make a careful analysis of the learning tasks in order to develop treatments and to create relevant aptitude measures. The analysis for this study and the resulting treatments are described in the next section. Afterwards, the creation of an aptitude measure relevant to these treatments is presented.

Treatments

Before designing the treatments the terminal objectives were specified. Then an analysis was made of the process of measuring area and coordinated with DMP's approach of spiraling through physical, pictorial, and symbolic representations. The result was a flow chart of behaviors as shown in Figure 2.1. These behaviors were later organized into instructional objectives (see Chapter IV).

The terminal objectives were:

1) Given a region and a covering of the region, the student

assigns a measure to the area of that region (see box 12 in Figure 2.1).

2) Given the measures of the areas of two regions, the student compares and orders the regions (see box 13 in Figure 2.1).

The two objectives in boxes 10 and 11 are those concerning the unit. The objective in box 10, identifies the need for knowing both the unit and the number, requires that the child focus on both the unit and the number. If he is always presented situations in which the number cue is correct then there is little need for him to focus on both. The objective in box 11 is closely related to the one in box 10. If two areas are being compared the comparison is simpler if each region has been covered with congruent units and these units are the same for both regions. Then, and only then, is it sufficient to compare only the measures.

This study hypothesized that some second and third graders do not need to be taught these objectives, some are ready to learn about them and others are not ready to assimilate these behaviors. Thus, the treatments were designed to either expose (Treatment U) or not expose (Treatment N) the subjects to these behaviors. Otherwise the treatments were held as constant as possible--the same length of time, the two treatment groups were randomly selected from a given population, the same teacher, the same instructional mode, and the same

terminal objectives.

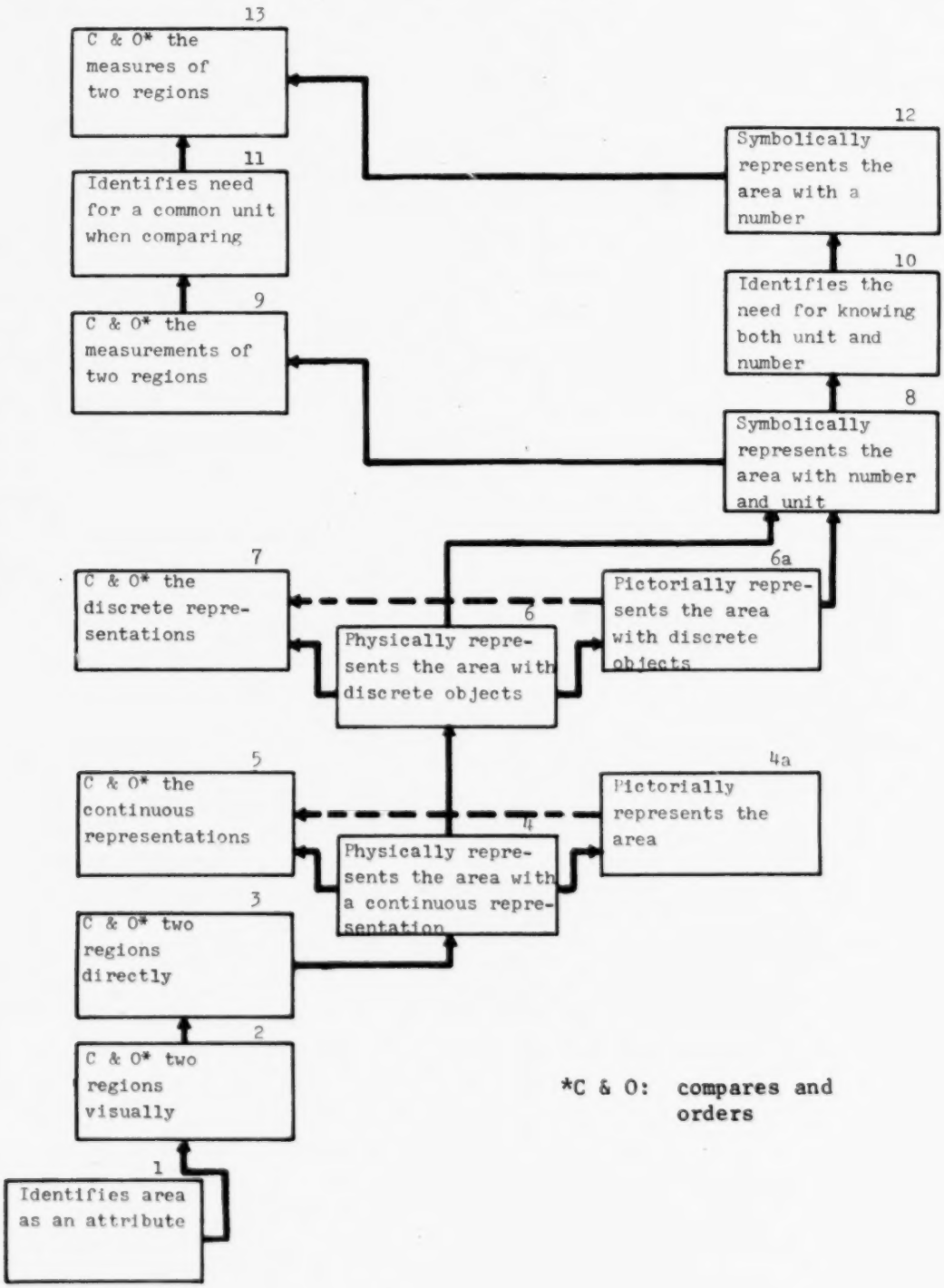
Aptitudes

Aptitude may be defined as "any characteristic of the individual that increases (or impairs) his probability of success in a given treatment." (Cronbach and Snow, p. 7) The characteristic of the individual identified in this study was the ability to learn about a unit of measurement. In order to determine this ability a teach-test procedure was used. This procedure consists of a pretest, a brief instructional unit, and a posttest. The underlying assumption of this procedure is stated in the following:

> This procedure is based on the supposition that if a student is unfamiliar with the content of the instructional unit, and if it can be reasonably assumed that he already has the background necessary for learning it, then his performance on the unit test will provide valid measure of his ability to learn mathematics; this measure could, in turn, be interpreted to be a valid measure of his mathematical aptitude. (Heimer and Lottes, p. 1-2)

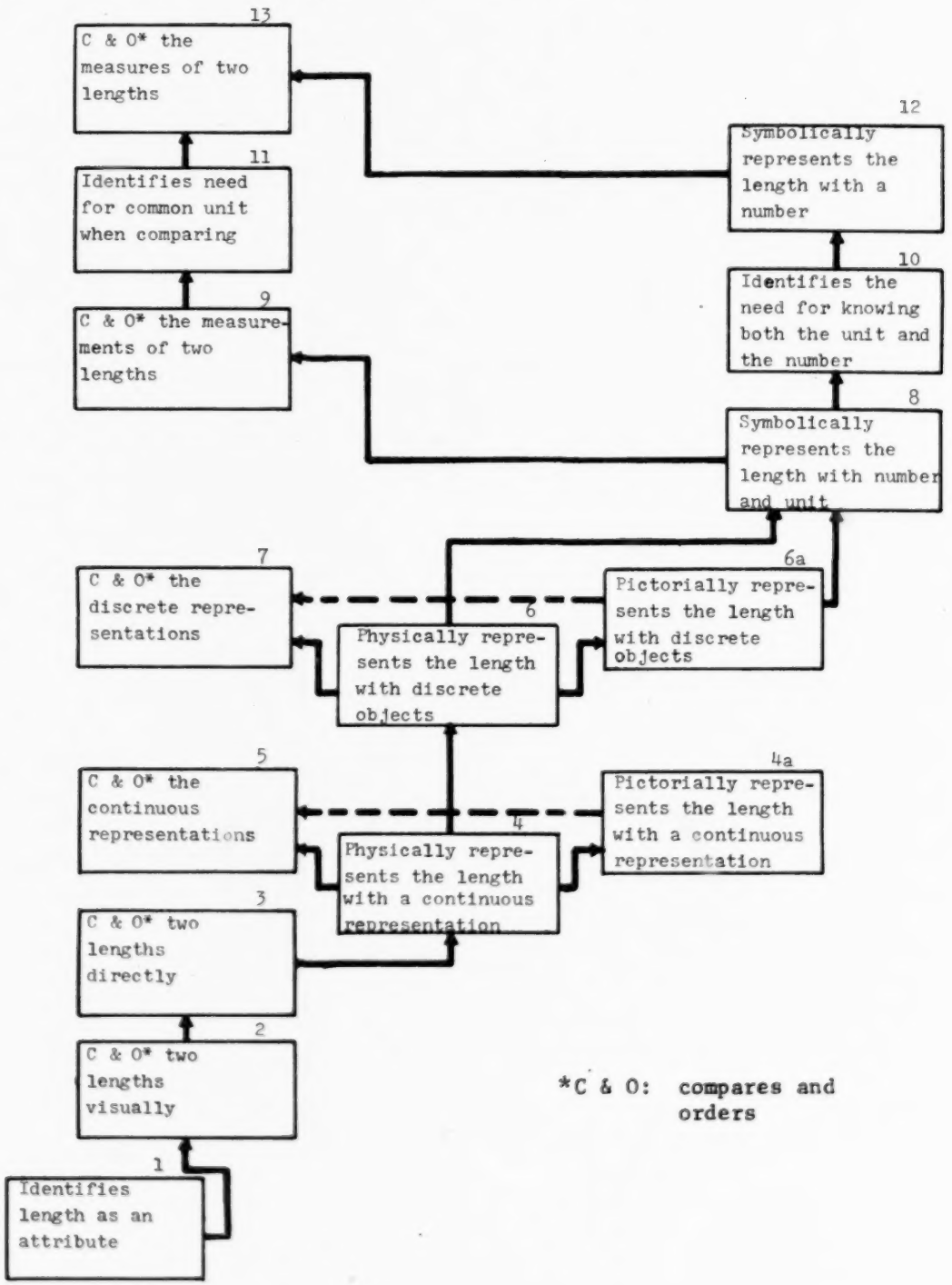
Since the instructional unit for the teach-test procedure is brief, it is necessary to limit the objectives of this instruction. To help accomplish this the number of attributes was limited in this study. It was decided to use only one attribute. The attribute of area could not be used because any instruction concerning the unit of area would interfere with the differences between the two treatments. The attribute of length was chosen for several reasons.

Figure 2.1. Flowchart for the Attribute of Area



*C & O: compares and orders

Figure 2.2. Flowchart for the Attribute of Length


13
C & O* the measures of two lengths
12
Symbolically represents the length with a number
11
Identifies need for common unit when comparing
10
Identifies the need for knowing both the unit and the number
9
C & O* the measurements of two lengths
8
Symbolically represents the length with number and unit
7
C & O* the discrete representations
6a
Pictorially represents the length with discrete objects
6
Physically represents the length with discrete objects
5
C & O* the continuous representations
4a
Pictorially represents the length with a continuous representation
4
Physically represents the length with a continuous representation
3
C & O* two lengths directly
2
C & O* two lengths visually
1
Identifies length as an attribute
*C & O: compares and orders

As indicated in the discussion of the process of measuring, measuring length and area involve the same steps. A flow chart for length (see Figure 2.2, derived in the same way as the one for area) shows the relation between the two attributes with respect to the behaviors under question.

While the steps to reach the terminal objectives (Figure 2.2, box 13) of comparing and ordering measures of length are the same as those for area, the behavior necessary for some of the subordinate objectives differ. The main differences are between the behaviors indicated in boxes 1 and 6 in both flow charts. Length is a much more easily identified attribute than area for children (see box 1 in Figures 2.1 and 2.2). Likewise, physical representation with discrete units is a much simpler procedure for length than for area (see box 6 in Figures 2.1 and 2.2). Thus, there is a strong logical and instructional relation between length and area, but the expected behaviors are easier to attain for length.

Furthermore, most second and third graders have attained many of the behaviors indicated in Figure 2.2. Prerequisite behaviors to the objectives concerning the unit are more likely to have been attained for length than for any other attribute.

After two tryouts of the teach-test procedure these objectives were specified:

1) Given two lengths, the student indicates that he must know both the number and the unit before he can compare

and order the lengths (see box 10 in Figure 2.2).

2) Given two lengths whose measurements have been expressed in different units, the student compares and orders these lengths by taking into account the relationship between the units (see box 11 in Figure 2.2).

These objectives were chosen for the following reasons:

1) They were found to be feasible from the pilot study.
2) They require only prerequisite behaviors which most second and third graders have mastered.
3) They are not stressed in elementary school programs and, hence, the interaction between the teach-test instructional unit and previous instruction would be minimal.
4) They are essential to the understanding of assigning measures to length and of comparing and ordering measures of length.
5) They are objectives specific to the unit.
6) They correspond to the differences between the treatments. Treatment U emphasizes the unit of area in a manner similar to the teach-test treatment of those objectives and Treatment N does not make this emphasis.

The teach-test procedure was used in the following way to classify students into aptitude levels. Level I: Any student who had not attained the specified behaviors by the end of the teach-test procedure and who had not evidenced change in performance. Level II: Any student who attained the specified behaviors only after the teach-

test procedure and who evidenced change in performance. Level III: Any student who had attained the specified behaviors before the teach-test procedure and maintained them throughout.

Specific Questions

In light of the definition of aptitude in this study the specific problem may be restated: How does children's ability, as determined by a teach-test procedure, to learn concepts about a unit of length interact with the two treatments on measuring area? Three dependent variables, achievement, transfer and retention, were identified. Thus, the question of interaction was asked with respect to each dependent measure.

In addition to these questions of interaction, several other questions were asked:

Is the teach-test procedure a valid predictor of an individual's success?

Is it feasible to teach these area concepts to second and third graders?

Is it feasible to teach about a unit of area or a unit of length to these students?

To what extent are the performances on achievement correlated with performances on retention?

Chapter III

RELEVANT AND RELATED RESEARCH

Introduction

While reviewing the educational and psychological literature one begins to feel that one is attempting to prove the following theorem:

> the amount of research in measurement is inversely proportional to the amount of use of measurement and support of its use in elementary school.

The amount of research is reflected in Suydam and Riedesel's (1969) interpretative study of research in elementary school mathematics in which only three of the reported 305 studies dealt solely with measurement. Only four more were cross-referenced with measurement. Although there is an abundance of pre-measurement research such as Piagetian conservation research, there has been little attempt to relate it to instruction. In one of the few attempts to relate Piagetian pre-measurement and measurement research to instruction, Huntington (1970) analyzed the instructional sequence of linear measurement in School Mathematics Study Group, Book 1. He found many discrepancies between this sequence and Piaget's developmental stages. Even if such discrepancies exist it is not clear at this time how curriculum developers could best use Piagetian research. Weaver (1972) in an article concerning the relevance of Piagetian research to instruction cautions mathematics educators:

> How significant it is that we have strained so hard to generate implications for classroom instruction in mathematics from research that thus far - with relatively few exceptions - could not have cared less for the classroom context and for the learning of mathematics by children within that context (p. 269).

Likewise, in reporting what research says about the use of measurement in other subjects, Suydam and Riedesel (1969) summarized:

> All the researchers agreed that greater emphasis should be placed upon understanding of basic quantitative concepts taught in elementary school mathematics (p. 117).

Although the research related to measurement instruction in proportion to its use and support is scarce, there are three branches of this research which are particularly relevant to this study. These are reviewed first in this chapter; afterwards the research related to aptitude-treatment interaction and that related to the teach-test procedure is reviewed.

Relevant Measurement Literature

The three branches of measurement research directly relevant to this study are research concerning: the role of the unit, the relation between measuring length and measuring area, and the measuring of area. Each of these is reviewed here.

The role of the unit. Ellison (1972, p. 171), in one of the few articles addressed to the role of the unit, warns of potential trouble that may occur with a unit. The main thesis of his article is the confusion that may arise if the unit is thought of as an identity

and is not conceived of as subject to change within specific problems. In surveying the current elementary textbooks, the investigator found that these sources of difficulty are not approached directly. In fact, there is little emphasis on the unit. Many of the teacher manuals touch on the role of the unit in measurement, but the texts do little for the children other than provide congruent units, discuss appropriate units for specific problems, and teach ways to represent the attribute.

The unit has been treated by research as it has been by curriculum, that is, incidentally. Major exceptions to this are studies by Piaget (1964), by Gal'perin and Georgiev (1969), and by Carpenter (1971).

Piaget's levels of development of measurement depend greatly on the child's facility with the unit. He describes children in Stage I and Level IIA as lacking in two understandings:

> In the first place they are ignorant of the composition of parts which means that they cannot understand that the set of cards (units) taken together equals the total area B, and that a part of that set alone is equal to the total area A, so that A < B because the part < the whole. In the second place they cannot see that the sections taken together equal the intact whole (p. 294).

Level IIB is a period of change:

> we see the beginnings of a common measure, ..., the beginnings of transitivity because there is better conservation. In other words, the composition of parts within a whole and the composition of positions and change of position are more coordinated (p. 295).

He divides his third stage into two levels:

> At Level IIIA, children use their composite common term to compare A and B but in doing so they simply count all the cards as if they were equal units and ignore their inequality; but at Level IIIB they understand the notion of a unit and so they take the size of the measuring elements into account (p. 295-296).

Thus to Piaget, measurement presupposes the understanding of the unit. However, his classical experiments were highly structured and many notions about the unit were not examined. Some of these notions were involved in the curriculum developed by and the research of Gal'perin and Georgiev (1969).

They designed fourteen individual exercises involving measurement tasks for a study with sixty children in a Russian kindergarten. These exercises involved measuring with a unit which was sometimes made up of parts, realizing the need for a common unit when comparing two measurements or realizing the need for common units. They were designed so that the relationship between visual comparisons and comparisons made by measuring could be examined. A posttest revealed that kindergarten children taught by traditional Russian methodology showed no improvement on such tasks. Gal'perin and Georgiev blamed this lack of improvement on the treatment of a unit:

> forming the concept of a unit as an entity results in an orientation that does not allow for the application of the unit as a means for measuring and counting. Such an orientation leads to direct comparison and visually quantitative distinctions (p. 194).

Using the results of this study and an analysis of the existing teaching methods, they proposed a series of operations to lead to the formation of elementary number concepts. Central to this sequence

was the role of the unit. Sixty-eight lessons were developed to reflect this sequence of operations. These lessons were used with 50 children in the same kindergarten as the initial study. Pretest scores on the same exercises were not much different (nothing but percentages of correct responses to items is reported) from the initial group, but posttest scores revealed that almost every child successfully completed every task. Their interviewing techniques also showed that these children had gained many sophisticated (for kindergarten) number concepts.

Although their research showed that it is feasible to teach notions about the unit to six and seven year olds and that there may be much payoff with number concepts, more important to this study is their conception of the role of the unit in measurement and in forming elementary number concepts. They build a strong logical argument for making the unit central to the development of elementary mathematics at the same time pointing out pitfalls that must be avoided.

Carpenter (1971) in his attempt to resolve the conflicting views of Piaget and Gal'perin and Georgiev devised 18 tasks for children in kindergarten through second grade. Ten of these tasks involved measurement and the unit. Either visually different units, indistinguishably (visually) different units, or the same units were used to measure the liquids in two containers. Carpenter also varied the true relation between the two liquids (equal or not equal) and the initial and the final visual relation between the two liquids (not seen, visually correct or visually deceiving). He also varied the relationship between the unit and the measure in such a way that when the true relation between liquid o_1 and liquid o_2 was $o_1 > o_2$, the units u_1 and u_2

were chosen in one case to make the measures m_1 and m_2 equal, and in the other case to make $m_1 < m_2$. Combinations of these variations produced the ten tasks.

In this portion of his study he individually tested 129 first and second graders on subsets of these tasks. He concluded that "by the end of the first grade, virtually all students realize that the quantity that measured the most units must be the greatest." However,

> only 70% of the Ss tested were able to use measurement results if they were followed by conflicting visual ones. Only 59% of the Ss tested demonstrated any knowledge that variations in unit size affected measurement results, and as few as 40% of Ss were able to apply this knowledge to problems in which quantities were measured with different units. This figure dropped to 25% of the Ss when the larger unit was not visibly distinguishable, and only 6% of the Ss were able to use results of measurement operations to determine the larger unit when it was not visually apparent (p. 99).

In conclusion he recommends that

> If one is really concerned with mastery of measurement concepts with different units of measure, it would seem necessary to provide a wide range of experiences that help the child focus on more than one immediate dominant dimension. It is important for teachers and curriculum developers to know when they are providing experiences that can be mastered, when they are providing experiences that may be learned superficially and when they are providing experiences that may be beyond the capabilities of many of the children (p. 106).

This investigator's study is partially addressed to gaining this knowledge - in reference to area measurement.

The relation between measuring length and measuring area. Several reasons were given in Chapter II for selecting the attribute of length for the teach-test procedure. One of these was that it was a simpler

attribute than area to measure. Logically and mathematically area is a more complicated attribute. There is overwhelming evidence that pedagogically length has been assumed to be the simpler. In all the textbooks reviewed measuring area never precedes measuring length. Paige and Jennings (1967) surveyed 39 series and found that most textbooks had introduced the standard unit of length by the third grade, but area measurement is not often introduced until fourth, fifth or even sixth grade (p. 356). However, there are few psychological studies which lend evidence to support or to deny this assumption.

Piaget (1964) proposes the same stagewise development for measurement of both length and area. In replicating his experiments Lovell, Healy and Rowland (1962) found the same stagewise development, but their research showed that the length stages are reached at an earlier age than the area stages. Beilin and Franklin (1962) showed that in training 6 and 8 year olds in length and area the older Ss could be taught both, but the younger could make progress only in acquiring length measurement.

Although no study could be found which gave direct evidence that understanding length measurement would facilitate the understanding of area measurement, research and practice does indicate that the converse is true. Hence, the investigator felt comfortable in hypothesizing that any subject who could not handle length measurement would have difficulty with area measurement. The other hypotheses concerning the interaction of students' ability to handle length and area were, admittedly, more open to question.

The measuring of area.

> The approach to area needs very careful consideration. In 1941 The Scottish Council for Educational Research published the results of their research into the teaching of area. Out of 1000 children 444 decided that the definition of area was "the distance all around it." Even today they make similar replies. The confusion seems to be due to ... (Marsh, 1969).

This quote represents a multitude of statements by individual educators and commissions. However, there are few studies other than conservation studies and Piagetian measurement studies (Beilin, 1964; Lovell, 1971; Needleman, 1970) concerned with determining the nature of this confusion. Beilin (1966) presented first and second graders with what he classified as quasiconservation area tasks. In contrast to the usual conservation tasks in which the transformation of one region is shown, he presented two regions, one of which had already been transformed. These tasks were actually measurement tasks in that the unit was specified and the subject was allowed to make the comparison between the two regions either by counting (termed iteration by Beilin) or by translocating (mentally moving the units of one region so that the shapes of the two regions are comparable). Several procedures were devised to train the subjects in iteration or translocation. Another experimental group received no training other than feedback of the correctness of their decisions on the pretest. Beilin found the feedback method was the most effective in producing change; thus he concluded that many subjects only needed to be reoriented to process the perceptual data differently. Beilin's study is relevant to this investigator's study for three reasons. First, he found that

second graders were more successful in area tasks than first graders. This gives reason for placement of area study in the second grade rather than the first. Second, he found that many more students naturally use the translocation method. This seems to indicate that the use of concrete materials which allows the student to physically make the translocations probably precedes merely counting static units. Third, he found that if students were reoriented to handle the perceptual data differently many could perform the tasks successfully. Hence, many students at this age seem only to need help in understanding the problem posed; it is not a problem beyond their ability.

Wagman (1969) hypothesized four levels of development of conservation of area similar to Piaget's. She found that children at all three age levels, 8, 10 and 11, had attained conservation and recommended that the study of area concepts should begin in the second or third grade.

The investigator found only two relevant studies which involved instruction in area concepts other than those involved related to the conservation of area. Luchins' (1949) descriptive study dealt mainly with the feasibility of teaching methods of, rather than formulas for, finding the area of triangles, rectangles and parallelograms to sixth graders. He also found that some five year olds could handle such problems. Luchins' claim that the intuitive approach to finding areas was more meaningful than a formula approach supports DMP's approach and hence, the approach to area adopted by this investigator. Johnson's (1970) study dealt mainly with the effect of varying the amount of

concrete experience in teaching perimeter, area and volume formulas to fourth, fifth and sixth graders. He found that those exposed to his maximum treatment of concrete materials were the most successful. He also found that age was a significant factor in learning concepts related to area. Children older than 11 years were more successful than those younger. His study lends support to this investigator's decisions not to introduce formulas at the second and third grade level and to rely heavily on concrete materials.

The well acknowledged difficulty which children encounter in area measurement combined with the paucity of research identifies this problem as one which requires major investigations.

Aptitude-treatment Interaction Literature

Since Cronbach's 1957 presidential address to the American Psychological Association in which he intimated that treatments may interact with abilities or personality traits, there has developed a substantial interest in aptitude-treatment interaction research. Cronbach and Snow (1969) in a comprehensive report on individual differences reviewed and critiqued many ATI studies and studies which lend themselves to ATI analysis. They concluded that most ATI studies have failed to produce interaction and contributed this to conceptualization problems, inappropriate analyses and the early state of the art. Although results to date have been discouraging, they urged against a defeatist attitude and suggested ways to attack some of the problems.

Bracht (1969) categorized 90 ATI studies according to the type of treatment, aptitude, dependent variable and interaction. In considering 108 of the possible interactions he found only five attained the goal of ATI, that is, had disordinal interactions. His definition of disordinal interaction is more stringent than Lubin's. Lubin (1961) only required that the regression lines of a significant interaction cross within the range of aptitude. Bracht added the restriction that the differences between the alternative treatments for at least two levels of aptitude must be significantly non-zero and must differ in algebraic sign. None of the five studies which he classified as disordinal were conducted in a classroom environment. The treatments were brief and controlled; the aptitudes included personality traits, abilities and social class; and the dependent variables were very specific tasks.

Subjects of all ages were used in the 90 studies reported. The number of subjects varied greatly from small studies of 30-40 subjects to ones with 700 or more. The five studies with disordinal interactions had a similar age span, but the average number of subjects was about 100. Thus, from this summary this investigator gained no direction as to age or to number of subjects.

A controlled treatment was defined by Bracht to be one in which "the degree of attainment of the treatment objectives was largely controlled by the presentation of specific and prescribed treatment tasks and little opportunity existed for the subjects to be influenced by other external conditions" (Bracht, 1970, p. 628). Eighty-five of the 108 studies including all five with disordinal interactions were

controlled. Many of these involved basic learning tasks or relatively short self-instructional, often computer oriented, units. Mathematical concepts were often used in the basic learning tasks, but four controlled studies (Becker, 1967; Behr, 1970; Carry, 1968; and Davis, 1968) investigated the learning of mathematics. These were all self-instructional units for secondary students; therefore, no further review of them is included here.

Two uncontrolled treatments involved mathematical instruction of third and fourth graders. Lucow (1964) examined the interaction of three levels of ability (IQ) with two six week treatments of multiplication and division, one with Cuisenaire rods and one without them. Using gain scores of 254 third graders on a multiplication and division test as a criterion, no interaction was found. Anderson (1949) found some tendencies toward ordinal interaction in looking at drill versus meaningful methods of teaching arithmetic to fourth graders. However, the distinction between the two methods and the teacher variable of the 18 classes makes one skeptical of even the tendencies.

This analysis of the treatments indicated that it was probably best to control the treatments as much as possible, even in a classroom setting for primary children. A warning from Cronbach and Snow, "educational policy cannot be based on what the pupil does with his first encounter with an instructional style," influenced the length of the treatments in this study. A brief treatment like many of the ones reviewed did not appear to be appropriate. Otherwise, this literature on treatments had little relevance to this investigator's study.

Aptitudes, or personological variables as they are referred to by Bracht, were classified according to whether they were factorially simple or complex. "Measures of specific abilities, interests, attitudes, personality traits, and social, economic and educational status were classified as factorially simple" (Bracht, 1970, p. 284). Measures of cognitive ability and achievement were considered complex. Most of the studies used general abilities such as IQ, but many used previous achievement. Specific abilities determined by tests from Guilford's Structure-of-Intellect battery or other similar tests constituted most of the factorially simple variables. Personality measures and other similar factorially simple variables were not often used.

No study was found which defined aptitude as the measure of the ability to learn. The only study which established aptitude by considering a task similar to the treatment task was a paired-associate study by Davidson (1964). Davidson determined his two ability groups on the basis of a pre-experiment paired-associate task similar to one of his five treatments. He found no overall interaction, but there was a tendency toward a significant interaction between two of his treatments and the aptitude. The lack of cell size would not permit a post hoc analyses of this interaction.

Thus, the studies reviewed lent little direction to the development of the aptitude measure used in this study. This direction came from the teach-test literature which is reviewed in the next section.

The dependent variable was usually specific; in fact, it was usually immediate achievement. One of Bracht's recommendations for

further ATI investigations is to use other measures such as transfer and retention. This advice was followed in designing this study. Past ATI studies have not proven to be successful in producing the desired interactions. However, they, along with critiques such as Cronbach and Snow (1969) and summaries such as Bracht, provide valuable assistance in planning further such studies.

Teach-test Literature

The motivation to use a teach-test procedure was derived from the work of Ralph T. Heimer. As far as the investigator could determine, Heimer and his associates' work (Heimer, 1966; Heimer and Lottes, 1968) constitutes the entire literature on this procedure.

Heimer describes in the American Mathematical Monthly, October 1966, the procedure and three studies. The first study was conducted with 106 entering college freshmen to ascertain the contribution this procedure could make with respect to predicting success in college mathematics courses. Although the results were inconclusive they showed that a high teach-test score corresponded to a high CMT (EST Cooperative Mathematics Test) but not vice versa. Concluding that "Teach-test apparently was discriminating among students at higher levels" (p. 885), Heimer was prompted in the summer of 1965 to conduct two parallel studies, one at Florida State University and the other at Florida A & M University.

Both of these studies were conducted in connection with the Secondary Scientific Training Program (SSTP), a summer program for talented

high school students. Twenty-three students at the beginning of the program at FSU were administered the teach-test package as well as Cooperative School and College Ability Test (SCAT), Form UA and the Cooperative Reading Test, Reading Comprehension, Form 1. At the end of the summer the instructors rated the students on a 14 point scale. Correlational analysis revealed that the T-T score was the highest correlate of the scale, although not significantly different from SCAT Quantitative score. It was an especially good predictor for the top students. The criterion used at Florida A & M was the rating of 1 to 25 of the twenty-five students involved. Again, teach-test procedure was good for predicting the exceptionally high students.

A more extensive study was conducted in the summer of 1967. Seven colleges offering SSTP courses participated in this tryout of teach-test packages. This involved 259 students and 14 courses. Beside the teach-test scores, scores on SCAT, II and IV (Insightful Computation and Mathematical Reasoning) and on five subtests of Personal Value Inventory (PVI) were attained. Two criterion scores, course grade and instructor's rating, were used in the analysis.

Five major hypotheses were investigated:

1. If the TEACH-TEST procedure is compared with conventional procedures for predicting scholastic success in the study of mathematics, then the TEACH-TEST correlation coefficient will be highest.

2. The TEACH-TEST procedure will measure on factors not taken into account by conventional procedures for predicting scholastic success.

3. The TEACH-TEST procedure will discriminate more effectively at the high-success levels than will

conventional procedures for predicting scholastic success.

4. If the TEACH-TEST procedure and conventional procedure for predicting scholastic success are compared on differences in effectiveness at different levels of prior educational opportunity, then the differences in effectiveness will increasingly favor TEACH-TEST as the level of prior educational opportunity decreases.

5. If the correspondence between TEACH-TEST content and criterion content increases, then the predictive effectiveness of TEACH-TEST will increase. (Heimer and Lottes, 1968, p 9)

Only the second hypothesis was supported by the analysis. However, there was some evidence to support the first hypothesis for some of the programs. There was not enough evidence to seriously test the fifth hypothesis, but there were trends which seemed to indicate its plausibility.

Thus, the most relevant information gained from these studies was the procedure itself. However, hypotheses 1, 2, 3, and 5 have some relation to the study or to recommendations for further study.

Although the first hypothesis was not supported there were some indications that if the course was closely related to the teach-test instructional unit the hypothesis was more plausible. The teach-test instructional unit for this study was constructed to be closely related to the treatments.

The second hypothesis indicates that the teach-test procedure measures factors not taken into account by conventional predictors. An interesting question to be pursued is whether a combination of such predictors would be a more valid criterion for determining aptitudes.

Heimer strongly suggests in all his studies that this procedure is more effective for predicting success of top students. By rejecting his hypothesis he failed to show that it was a more effective discriminator than conventional predictors at the top fifth level or the second fifth level. Although this investigator's study cannot test this hypothesis directly, its data could easily be used to test the hypothesis that the teach-test procedure was a better predictor for the upper level of aptitude than for the lower level of aptitude.

In the recommendations for further study, Heimer and Lottes make a plea for the tryouts of teach-test packages which closely correspond to the content of the course. This study provides the opportunity to reexamine their fifth hypothesis by testing whether the teach-test procedure was a better predictor for Treatment U than for Treatment N.

Although Heimer's work is inconclusive and there is danger in generalizing any of his findings, it did provide the motivation and guidelines for using the teach-test procedure in this study.

Summary

On the one hand, the search for relevant literature did not prove to be very fruitful or encouraging. The research on measuring was scanty and often not directly relevant to this study. The ATI research results were discouraging. The research utilizing the teach-test procedure was of a neophyte nature.

On the other hand, this very lack of research and of positive results coupled with the conviction of mathematics educators that

measurement is an important topic and the support of some prominent educational researchers for the pursuit of ATI studies, especially those which define aptitudes in new ways, gave encouragement to the investigator.

Thus, with whatever knowledge that could be gleaned from past research and experience, the investigator proceeded to plan the study, a description of which follows in the next chapter.

Chapter IV

DEVELOPMENT OF THE STUDY

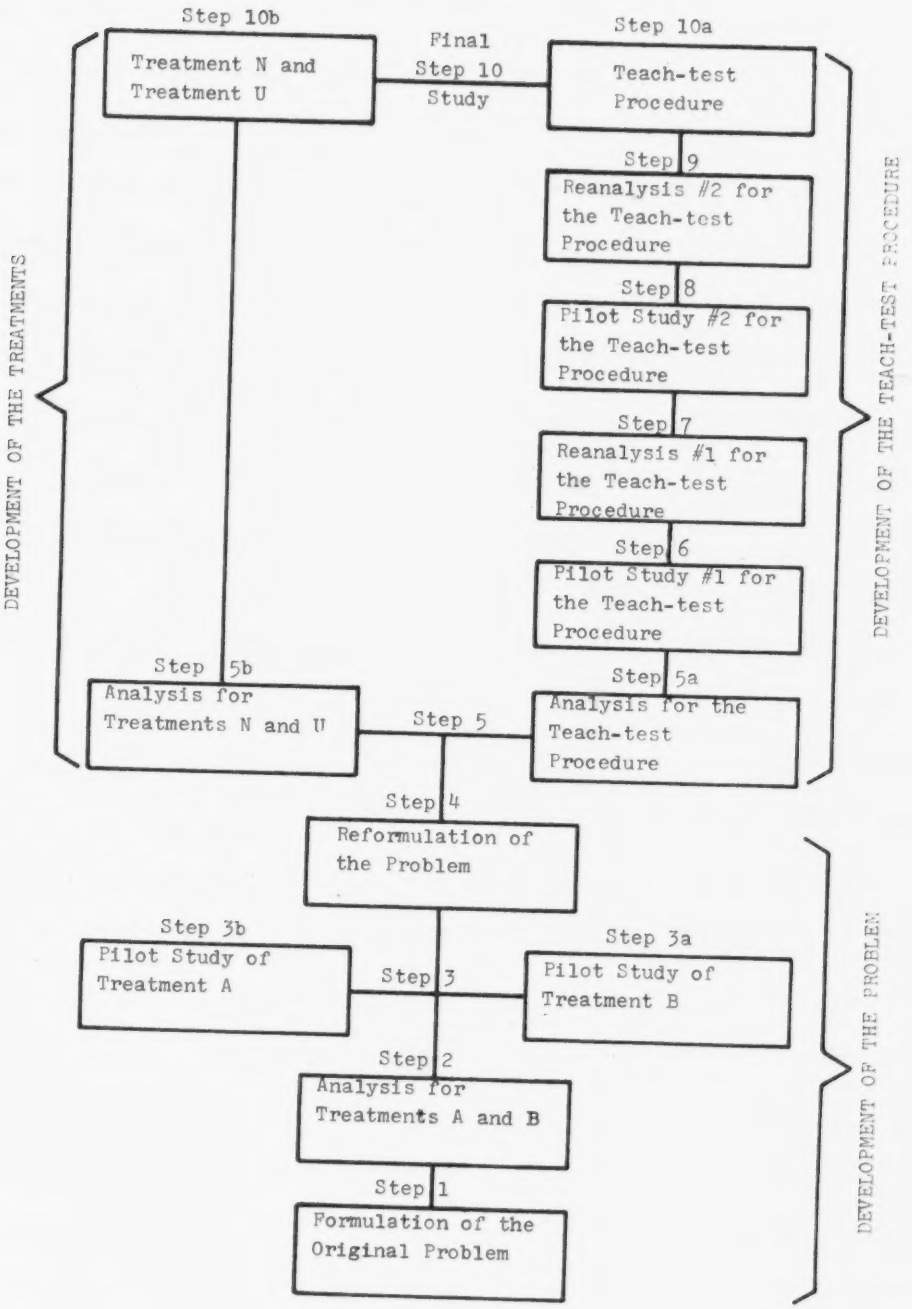
Introduction

This chapter gives a detailed description of the development of the study. First, the development of the experiment is detailed and then the research strategy used in conjunction with this experiment is described.

The instructional treatments, the teach-test treatment and the observation instruments used in this experiment were developed in accordance with the first two phases of the curriculum development model of Romberg and DeVault (1967). These phases, analysis and pilot, are iterative. That is, first both mathematical and instructional analyses are made of the curriculum to be developed. Then the curriculum is developed and piloted. After analyzing the pilot there are two choices: (1) if the results of the pilot were satisfactory one proceeds to the next phase, validation, and (2) if the results are not satisfactory one recycles through the first two phases.

Figure 4.1 shows the cycle of analysis and pilot phases for this experiment. The steps have been grouped into three categories, development of the problem, development of the teach-test procedure and development of the treatments, to facilitate the discussion which follows.

Figure 4.1. Development of Experiment



Development of the Problem (Steps 1-4)

This section describes the development of the problem by describing the original problem, the pilot studies conducted to investigate this problem, and their influence both on the formulation of the final problem and on the treatments used in the final study.

Step 1: Formulation of the original problem. The formulation of the original problem was in response to the curriculum development described in Chapter II. Originally, the problem was to investigate the relative effectiveness of two instructional treatments of area; one reflected the DMP approach (Treatment B) and the other reflected a more traditional approach (Treatment A). No specific hypotheses were proposed; the purpose of the first pilot studies was to informally investigate this problem.

Step 2: Analysis for the Treatments A and B. Even for an informal investigation both a mathematical analysis of the content and an instructional analysis must be made. These analyses were made but are not reported here since the mathematical analysis is described in Chapter II and the instructional analysis is reflected in the description of the pilot studies which follows.

Step 3: Pilot studies of Treatments A and B. In the spring of 1971 both Treatment A and B were piloted with a small number of children at Randall Elementary School, one of the DMP developmental schools.

Treatment B was designed to reflect an attribute-by-process approach. That is, an attribute is identified and processes are extended or developed with that attribute. In this case, the attribute of area was

identified and the processes of comparing, ordering and measuring were introduced sequentially. Activities were developed which involved superimposing, exact and approximate coverings, cutting and rearranging. Nine 25-minute lessons and an individually administered test was developed. Six first graders from several classrooms participated; however, due to illness only four were present for the entire treatment. The small sample gave the investigator an opportunity to ask many probing questions. Although many of the activities were suitable for first graders it was felt that too many inferences were expected for these children.

Treatment A developed area in a way similar to many traditional textbook series. Area was defined as a number. The children were asked to assign a measure, by counting, to a region whose covering with units was shown. Using numbers assigned in this method, the children were asked to compare and order the regions on area. Thus, area was always considered as a number; no direct comparing, cutting or rearranging was ever required. Four second graders participated in this treatment which was planned for five days but was completed in three.

Although both groups were formally tested the information gained from the testing was considered as only another observation of the children and their responses. From the two treatments and tests many subjective decisions were reached:

First, the approach taken in Treatment B was preferable to that in Treatment A. The investigator felt that those in Treatment A were only manipulating numbers with no reference to area; this feeling was

confirmed by the test in which the children were asked first to visually compare pairs of regions and then to compare the same regions covered by pieces in such a way that the comparison of the number of pieces did not correspond to the comparison of the area. They were able to do the former but not the latter task.

Second, many instructional problems or successes were evident which needed to be taken into account when designing the treatments for the final study.

Third, first grade was too early to expect much success in a treatment which went as far as Treatment B; although many of the earlier activities are suitable for this age group. Thus, the study should be conducted with an older group of children.

Fourth, it was possible to design a group test for second graders; however, the format needed to be simplified.

Finally, the investigator felt a more interesting and important question became evident. In both groups the lack of understanding of the unit of area was a problem for some children but not for others. What could be done, if anything, for those children who lack this understanding? What is best for those children who already have this understanding?

Step 4: Reformulation of the problem. In summary, this pilot study not only gave the investigator more insight into the pedagogical problems, but also changed the direction of the study. Although the relative effectiveness of these two treatments should be investigated, both treatments depend upon the child's understanding of the unit.

Therefore, before looking at the contrast between these treatments the investigator decided to contrast two treatments by their emphasis on the unit. Using Treatment B as a basis, two treatments would be developed; one of which would stress the unit and the other which would not stress the unit. This involved reconceptualization of the problem and led to the formulation of the problem as described in full in Chapter II.

The question asked was: How do children's abilities to learn about a unit of length interact with two treatments on measuring area? A teach-test procedure was used in order to determine the children's ability to learn. The development of this procedure is described next. Afterwards the development of the treatments is given.

Development of the Teach-Test Procedure (Steps 5a - 10a)

The teach-test procedure and the treatments were planned together so that the desired interaction could be taken into account. However, for ease and clarity of description the development of the teach-test procedure is discussed first. As stated in Chapter II the decision was made to teach about the unit of length in the teach-test procedure. The rationale for that decision was presented there; this section describes the evolvement from the first pilot study of this procedure to the final form of this procedure.

Step 5a: Analysis for teach-test procedure. There appear to be two prominent misunderstandings about a unit of measure: the need to use congruent units when assigning a measure and the need to use common units when comparing two regions. The first pilot study of the teach-

test procedure focused on both of these needs. A test and two lessons were developed to test and teach the following objectives:

1. Given a length the student indicates that it must be represented with congruent units in order to assign a measure to it.
2. Given two lengths whose measurements have been expressed in terms of different units, the student indicates that the lengths cannot easily be compared unless the same unit is used to represent both lengths.

The first lesson presented the children with the problem of representing lengths with congruent units as a communication problem: "How do you tell someone else the length of an object if non-congruent units represent it?" After a group discussion the children worked in small groups and individually on similar problems.

The second lesson presented the problem of comparing two lengths when non-common units have been used to measure the lengths. Again the problem was posed in a communication context. In the second part, the children had a contest in which each measured an object with congruent units and then they tried to compare their length with an opponent's length which may have been measured by a different unit. In the third part they responded to similar comparison questions on a worksheet.

Step 6: Pilot Study #1 of the teach-test procedure. The investigator along with the teacher who participated in the final study tried this version of the teach-test procedure with 23 third graders at Randall Elementary School. On the first day a pretest was given and

Lesson 1 was covered. On the second day, Lesson 2 was covered. A post-test was scheduled for the end of the second day, but the second lesson lasted longer than the forty minutes allotted, so the posttest was given the third day.

Step 7: Reanalysis #1 for the teach-test procedure. Many observations were made during the first pilot study. The first lesson did not go well; the children did not see the need to represent the length with congruent units. They were perfectly happy to describe the representation more fully by saying, for example, three short units and four long units. In the next lesson when they were asked to compare lengths, they then saw a need for congruent units as well as common units. Thus, it was decided to de-emphasize the first objective, and to drop the first lesson. The contest in the second lesson created much enthusiasm but was unmanageable. It was decided to save such activities until the main treatment when the teacher knew the pupils better. Problems with worksheets of the second lesson were found. Many children were ignoring the measurements and relying on the visual stimuli, although the stimuli were often deceiving. Too much was planned to be covered in two days; this resulted in two changes, a shortened treatment and better organization so that no time would be wasted in management. The children appeared capable of handling simple ratio relationships between units. This capability was utilized more in the next pilot study. There were positive aspects of the two days which were retained in the subsequent tryout.

The test which was used for both the pretest and the posttest con-

sisted of 23 items. Three items tested prerequisite behaviors and the other items tested the two objectives. Two difficulties were noted with the test. First, it was too easy; this was partly due to always showing the objects whose lengths were to be compared. The lengths could always be compared visually or by using a pencil to represent one length and checking this against the other. Thus, there was no need to rely on the measurements. Second, the use of the same test for both the pretest and posttest caused both motivational problems ("I knew this before, why do I have to show you again.") and measurement problems ("I remembered I answered M on this before.") Otherwise, except for minor exceptions the test was clear and the length was appropriate. The results of the tests showed that most of the children had mastered the objectives both on the pretest and the posttest. The investigator felt this was an artifact of the test since the children did not have to rely on the measurements.

Step 8: Pilot Study #2 of the teach-test procedure. The teach-test procedure was modified in accordance with the findings of the first pilot and piloted again. This time, twenty-one second graders at Randall Elementary School participated. On the first day, after the pretest, the first lesson presented was comparing lengths that had been represented with different units. On the second day, the children worked more independently with problems of comparing and ordering lengths which were represented with different units.

Step 9: Reanalysis #2 for the teach-test procedure. Although the second pilot study went more smoothly, several changes were made for

the final study. The ambiguous questions on the worksheets were eliminated and the format simplified. The test items which were unclear and the items which required a "cannot tell" response were changed so that the comparison could be made. The second graders had some difficulty in following the complex directions and these changes enabled the directions to be much more straightforward.

The results of the pretest and posttest indicated a positive change for about half of the children over the treatment. A few children had mastered the objectives on the pretest; many more had on the posttest. The other half showed no change in performance.

Step 10a: The final plan for the teach-test procedure. As a result of the pilot studies and subsequent reanalyses the final teach-test procedure was designed.

Two objectives were specified:

1. Given two lengths the student indicates that he must know both the number and the unit before he can compare and order the lengths.
2. Given two lengths whose measurements have been expressed in different units, the student compares and orders the lengths by taking into account the relationshi between the units.

The lesson plans may be found in Appendix A. Essentially, the two lessons remained the same as they were for the second pilot. The first lesson was planned to last twenty minutes and to be given immediately following the pretest. It is a large group activity which

introduces the difficulty of comparing lengths that have been represented with different units. This is done by representing two lengths L_1 and L_2 with units U_1 and U_2, respectively, and with the measures M_1 and M_2, respectively, in the following ways:

1. Given $L_1 > L_2$,

   U_1 and U_2 are chosen so that $U_1 < U_2$ and $M_1 = M_2$,

2. Given $L_1 > L_2$,

   U_1 and U_2 are chosen so that $U_1 > U_2$ and $M_1 < M_2$,

   and

3. Given $L_1 > L_2$,

   U_1 and U_2 are chosen so that $U_1 = U_2$ and $M_1 > M_2$.

The second lesson was planned to last forty minutes. In the first half the children work individually comparing and ordering two lengths. Each child receives a packet of strips which are color coded with respect to length. These help him answer questions presented on a worksheet about the order of the lengths of, say, 5 red strips and 7 blue strips. The questions are structured so that the measures alone do not always specify the order relation between the two lengths. The second half of the lesson is designed as a large group activity in which the children are shown two measurements without the visual stimuli of the two lengths. They are asked to compare and order the lengths and to give reasons for their decisions. The two instruments O_1 and O_2 developed for the teach-test procedure may be found in Appendix B. The pretest O_1 consisted of 22 items; the first two items tested prerequisite behaviors. These items were used to eliminate any student

who did possess these behaviors, but not in any further analysis. The remaining twenty items tested the two objectives of the teach-test. The items varied the type of stimuli; combinations of the following possibilities were made. The two lengths were either visually comparable, visually misleading, or not shown; the units representing the two lengths were either the same or different; and the relation between measures was either in the same or different order as the actual lengths. Thus, the student could not depend upon visual comparison or numerical comparison, but was forced in most instances to coordinate the measure and the unit. The posttest O_2 consisted of items from the pretest arranged in a different sequence and denoted with different labels.

Development of Treatments (Steps 5b, 10b)

The two treatments were developed concurrently with each other and with the treatment for the teach-test procedure. There were two main steps in this development. First, an extensive analysis, both mathematical and instructional, was made for the treatments. The instructional analysis included a reanalysis of the pilot study of Treatment B. Secondly, the specific lesson plans and testing instruments were developed. This section describes both of these steps.

Step 5b: Analysis for the treatments. The first part of the analysis was a mathematical analysis of the process of measuring as it related to area. This analysis which was described in Chapter II resulted in the selection of the eight behavioral objectives listed in Table 4.1.

Table 4.1

BEHAVIORAL OBJECTIVES FOR TREATMENTS
U AND N

1. Given a region, the student identifies area as an attribute of that region.
2. Given two regions, the student visually compares and orders them on the attribute of area.
3. Given two regions, the student directly compares and orders them on the attribute of area.
4. Given a region, the student physically represents the area with discrete objects.
5. Given a region, the student symbolically represents the area with a number and a unit (measurement).
6. Given the measurements of the areas of two regions, the student compares and orders them using their measurements.
7. Given a region, the student symbolically represents the area with a number (measure).
8. Given the measures of the areas of two regions, the student compares and orders them using their measures.

Each treatment group was to be sequenced through these eight behaviors. However, some of the objectives are interpreted differently for the two treatment groups. The first three objectives do not ex-

plicitly involve a unit, thus they were treated the same across groups. Differences between the treatments are reflected in Objectives 4 and 5. A region may be represented either with congruent units or non-congruent units. Subjects in Treatment N represented regions only with congruent units, but subjects in Treatment U also used non-congruent units. Thus, subjects in Treatment N had less reason than those in Treatment U to record the unit used in representing. Objective 6 was also treated differently across groups. Treatment N's subjects were presented regions which they covered or which were already covered with common units, i.e., both regions were covered with the same unit. Treatment U's subjects were asked to compare and order regions which were not covered with common units. Thus, for Treatment N's subjects there was little reason to focus on the unit when comparing, for the number would always suffice. Both treatments were to teach toward objectives 7 and 8 through objectives 5 and 6, respectively. Subjects in both treatments were presented with the same types of simple regions and the same type of units. All measures in both groups were assigned only by counting; no formulas were introduced.

The next step in the analysis of the treatments was the instructional analysis. In this process the interactions of the learner with the mathematical content must be carefully considered. Since there are no theories of instruction which lend adequate guidance to this process, the investigator resorted to the knowledge gained from the pilot studies of Treatments A and B and to the expertise of those who have developed instructional activities for DMP.

Several decisions were reached common to both treatments. First, both treatments were to reflect the instructional mode of DMP. That is, both would center on an activity approach which uses physical materials and problem solving whenever feasible. Secondly, the sequencing of activities would be similar. The first three days would be spent acquainting the students with the attribute of area, with the units used in measuring and with activity learning. The next three days would be spent mainly on representing area and the last three days on comparing and ordering areas using their measurements or measures.

Next, each behavioral objective was further analyzed in terms of pedagogical considerations. Instructional objectives for each treatment were written which taught toward the behavioral objectives. Table 4.2 lists these objectives and keys them to the behavioral objectives listed in Table 4.1.

Step 10b: The final plan for the treatments. From these instructional objectives nine days of activities were planned for each group. Each day's session was to be forty minutes long; because flexibility was desirable more activities than were considered necessary for each forty minutes were planned. Each day's plan included the behavioral objectives, the instructional objectives, the materials, the organization and a description of the activity. These, as well as the children's activity sheets and a daily journal, are found in Appendix C. The purpose of each day's activities and the contrast between treat-

Table 4.2

INSTRUCTIONAL OBJECTIVES FOR TREATMENTS U AND N

Treatment U			Treatment N		
Instructional Objectives		Behavioral Objectives	Instructional Objectives		Behavioral Objectives
U.1:1[a]	To introduce the attribute of area.	1	N.1:1	To introduce the attribute of area.	1
U.1:2	To compare and order regions visually on area.	2	N.1:2	To compare and order regions visually on area.	2
U.1:3	To introduce superposition as a method of comparing and ordering regions on area.	3	N.1:3	To introduce superposition as a method of comparing and ordering regions on area.	3
U.2:1*	To see that a region consists of parts.	4	N.2:1*	To see that a region may consist of congruent parts.	4
U.2:2	To give experience with the shapes which are to be used as units of area.	4	N.2:2	To give experience with the shapes which are to be used as units of area.	4
U.2:3*	To introduce the relationships between the units of area.	5,6			
U.3:1*	To introduce physically representing area as a method of comparing and ordering regions.	4,6	N.3:1*	To introduce physically representing area with congruent units as a method of comparing and ordering regions.	4,6

Table 4.2 (continued)

Treatment U		Treatment N	
Instructional Objectives	Behavioral Objectives	Instructional Objectives	Behavioral Objectives
U.3:2* To consider measurements of the area of an object represented by different units.	4	N.3:2* To practice physically representing with congruent units the area of objects in the room.	4
U.4:1* To practice physically representing the area of a region.	4	N.4:1* To practice physically representing the area of a region with congruent units.	4
U.4:2* To introduce symbolically representing the area of a region which has been covered with either congruent or non-congruent units.	5,7	N.4:2* To introduce symbolically representing the area of a region which has been covered with congruent units.	5,7
U.5:1* To introduce choosing an appropriate unit to represent the area of a region.	4	N.5:1* To introduce regions which cannot be completely covered with a given unit.	4
U.5:2 To practice symbolically representing the area of a region.	5,7	N.5:2 To practice symbolically representing the area of a region.	5,7
U.6:1* To introduce regions which cannot be completely covered with a given unit or a mixture of units.	4	N.6:1* To practice assigning measurements to regions which cannot be covered completely with a given unit.	5,7

Table 4.2 (continued)

Treatment U			Treatment N		
	Instructional Objectives	Behavioral Objectives		Instructional Objectives	Behavioral Objectives
U.6:2	To practice symbolically representing the area of a region.	5,7			
U.7:1*	To compare and order regions whose area measurements are expressed in terms of either common or non-common units.	6,8	N.7:1*	To compare and order regions whose area measurements are expressed in terms of a common unit.	6,8
U.8:1*	To practice approximating the area of a region when it is not covered exactly.	5,7	N.8:1*	To practice comparing and ordering the areas of two regions on area by using 1/2" square grids.	6,8
U.8:2*	To compare and order regions on area whose measurements have been approximated.	6,8			
U.9:1	To practice comparing and ordering regions on area using their measurements.	6,8	N.9:1	To practice comparing and ordering regions on area using their measurements.	6,8

a) U.1:1 refers to the Treatment U, day 1, instructional objective 1; similar notation.

*) Indicates that the objective was not common to both treatments.

ments are summarized here.

Day 1: The purposes and the activities for both treatments are the same. Besides the three instructional objectives (to introduce the attribute of area, to compare and order area visually, and to introduce superposition as a method of comparing and ordering regions) the purposes of this lesson are to have the children verbalize as much as possible and to have the children begin solving problems on their own. For example, after comparing regions visually, the children are presented two regions which may be visually deceptive to them. They are to find a way to compare the regions or to verify their decision about the comparison.

Day 2: Although one purpose of this lesson is to acquaint the children with the plastic pieces they are to use for units, it is expected some would see that a region is made up of many parts and that the same pieces make many different regions.

In Treatment N the children use congruent pieces to make a larger region and transfer the region to paper tesselated with that piece. In Treatment U the children also use non-congruent units to make a larger region and transfer the region to paper tesselated with one of the units. For example, they make a design with squares and right triangles and transfer the design to a paper tesselated with squares. Or they make a design with rhombuses and transfer it to equilateral paper.

Day 3: Both groups are given the problem of comparing the area of two objects in the room which cannot be superimposed. Through this problem the need for physically representing the area by covering is

introduced. In Treatment N the children find the area of objects in the room by representing them with congruent units. The children in Treatment U do the same activity except that two different measurements of each object are found. The procedure of covering is refined in subsequent activities.

Day 4: The purpose of this activity is to practice representing regions with discrete units and to assign measurements to the regions. In Treatment N the children cover only with congruent units and write the number of units required to cover the region. In Treatment U the children also cover with non-congruent units and write the measurement in terms of all units used or they look at the relationship between the units and write the measurement in terms of one unit. Both groups are presented regions which may be covered exactly or which are covered exactly with a given unit.

Day 5: Both treatment groups are presented with regions that cannot be completely covered by a given unit. However, in Treatment U the children recover with a unit or units that cover exactly.

Day 6: Again both treatments receive practice in assigning measurements to regions which cannot be covered exactly and in estimating areas. Treatment U's students cover the same region with different units. In so doing they are led to see that a smaller unit may cover more exactly. Treatment N's students cover the same regions but only use one unit.

Day 7: Both treatments compare and order regions on the basis of their measurements. In Treatment N the coverings are presented or the

children cover both regions with common units. In Treatment U the coverings may be with non-common units, thus the children must take into account the relationships between the units to make the comparisons.

Day 8: Again the basic purpose of this day's activities is to compare and order areas. Treatment N introduces a one-half square inch grid to assist in making the comparisons. Treatment U emphasizes the need to approximate areas of regions that cannot be covered exactly.

Day 9: This day is a summary of previous activities in which both groups receive additional practice in comparing and ordering areas.

To complete the development of the treatments instruments were constructed by the investigator to measure the three dependent variables. These instruments may be found in Appendix D.

The same instrument was constructed for both the achievement measure (O_3) and the retention measure (O_5). It consisted of 30 items; six tested assigning measurements to regions, two tested visual ordering of regions, four tested ordering regions when only the measurements were given, and the remainder tested ordering two regions after assigning a measurement to each. Out of the 22 items which tested comparing and ordering with measurements, eight were covered with congruent units. The remaining 14 required coordination of both the unit and the measure, or expressing the measurement of the region in terms of one unit. Regions which were covered exactly and those which were not covered exactly were used as stimuli.

The transfer measure (O_4) consisted of 25 items. Ten of the items involved area and five involved length. These fifteen items all asked

questions about the unit which were more complicated than those of the teach-test procedure or the treatments. The remaining ten items were divided between the attribute of capacity and of numerousness. All questions dealt directly with the unit of measurement. Thus, the transfer test was an extremely heterogeneous one.

The Research Strategy

This section includes the research design used in the study, the population, the hypotheses and the proposed statistical analysis.

Research design. As stated in Chapter II the problem was posed in terms of an aptitude-treatment interaction question. The basic research design for an aptitude-treatment interaction study is one in which the identified aptitude is measured, the subjects are assigned to a treatment, and the outcome is measured (Cronbach and Snow, 1969, p. 21). This design was adapted for this study in the following manner:

1. The identified aptitude is measured by a teach-test procedure which includes a pretest (O_1), a brief instructional unit (T-T), and a posttest (O_2). Three levels of aptitude, Levels I, II, and III, are determined, according to the following specifications: Level I consists of those students who are at the nonmastery level on both tests. Level II consists of those students who have not attained mastery on the pretest but have attained mastery on the post-test. Level III consists of those students who are at the mastery levels on both O_1 and O_2.

2. Subjects from each of Levels I, II, and III are randomly assigned either to Instructional Treatment N (N) or to Instructional Treatment U (U) or they are randomly excluded from further treatment. This is done in such a way that each instructional group is assigned m, n, and p students from Levels I, II, and III, respectively. Subjects who do not fit either Level I, II, or III, i.e., those who attained mastery on O_1 and not on O_2 are to be excluded.
3. The outcomes include immediate achievement and transfer measures from observations O_3 and O_4, respectively, and retention measures from O_5.

The design is summarized in Figure 4.2.

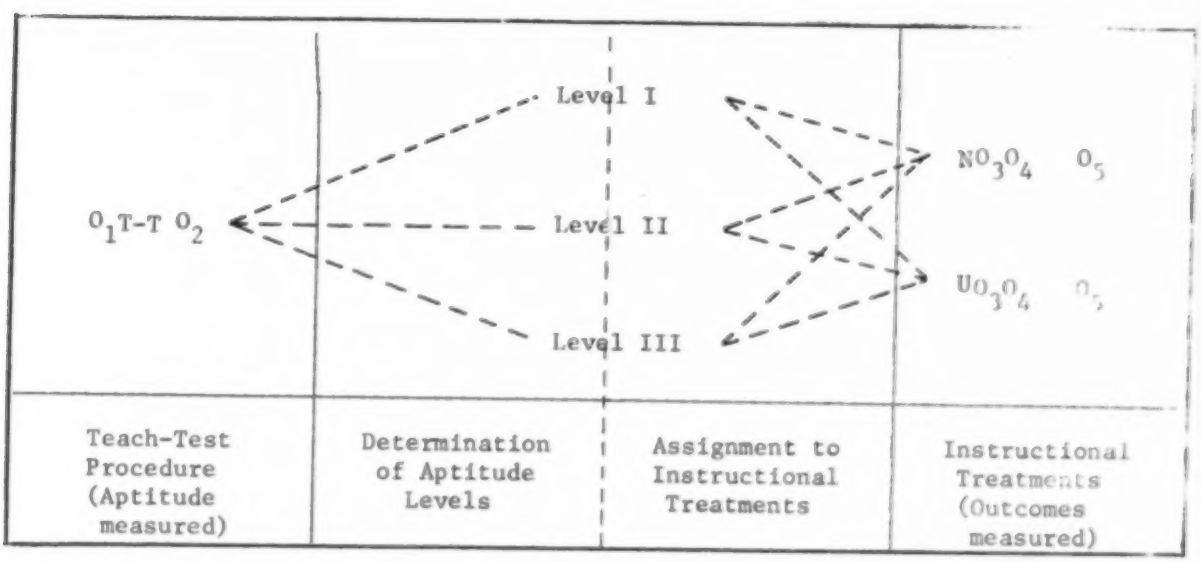


Figure 4.2. Research Design

Thus, the independent variables are the teach-test treatment (TT) and the instructional treatments (N and U). The dependent variables are the observations O_1 - O_5. Observations O_1 and O_2 are two forms of the same instrument which differ only in the order of questions and labels of stimuli. One instrument is used for O_3 which tests immediate achievement and another instrument is used for O_4 which tests transfer. Observation O_5 which tests retention is made with the same instrument used for O_3. Both observations O_3 and O_4 will be made immediately following the treatments and observation O_5 will be made approximately four weeks later.

Population. From the pilot study conducted in the spring of 1971 it was concluded that the second or third grade was the proper placement for the proposed experiment. In order that all levels of the defined aptitude be represented in the instructional treatments, it is recommended that the teach-test treatment include three times more subjects than each of the instructional treatments. Thus, if each instructional treatment has 30 subjects, 90 subjects would be needed for the experiment.

Hypotheses. For each dependent variable, achievement, transfer and retention, three research hypotheses were posed. In each case one of these three was of primary concern in this study and is designated the primary research hypothesis while the other two were of a secondary nature. A primary research hypothesis refers to the

interaction of the aptitude with the treatment while a secondary research hypothesis refers to the main effect of either aptitude or treatment. Thus, the following nine research hypotheses were posed:

Primary research hypothesis 1a. There is a significant interaction between the ability to learn about a unit of length as determined by the teach-test procedure and the two treatments when measured by achievement.

Secondary research hypothesis 1b. There is a significant difference between the levels of aptitude on their achievement performance.

Secondary research hypothesis 1c. There is a significant difference between the treatment groups on their achievement performance.

Primary research hypothesis 2a. There is a significant interaction between the ability to learn about a unit of length as determined by the teach-test procedure and the two treatments when measured by transfer.

Secondary research hypothesis 2b. There is a significant difference between the levels of aptitude on their transfer performance.

Secondary research hypothesis 2c. There is a significant difference between the treatment groups on their transfer performance.

Primary research hypothesis 3a. There is a significant interaction between the ability to learn about a unit of length as determined by the teach-test procedure and the two treatments when measured by retention.

Secondary research hypothesis 3b. There is a significant difference between the levels of aptitude on their retention performance.

Secondary research hypothesis 3c. There is a significant difference between the treatment groups on their retention performance.

The arbitrary decision was made to reject the corresponding null hypotheses at the .05 level of significance.

Proposed Statistical Analysis. Although there exist more sophisticated techniques for testing the presence of an aptitude-treatment interaction (Cronbach and Snow, 1969, p. 23), there are two common

techniques: analysis of variance and regression analysis. In this study both of these common techniques would produce the same F-ratios (Kelly, Beggs and McNeil, 1969). This study does not lend itself to more sophisticated statistical techniques, so that a 3 x 2 ANOVA is proposed (three levels of aptitude by two treatments).

There are several assumptions which must be met in order to justify using an analysis of variance (Hays, 1965, p. 396). First, the errors must be normally distributed for each of the aptitude-treatment populations. This assumption may be violated provided the number of subjects is relatively large. This and the random assignment to treatment groups should help assure that the F-test would be unaffected by non-normality. Second, the errors must have the same variance for each aptitude-treatment group. This assumption may be violated without serious consequences if the number of cases in each example is the same. Thus, it is planned to have equal or as near to equal as possible all cells. Third, and the most important assumption, is that of independence of observations both across and within treatment groups. The interaction across treatment groups is planned to be held to a minimum, but the use of individual scores as a basis of analysis does not meet the assumption of independence within treatment groups. The investigator was well aware of the violation of this assumption. However, it was impossible to secure a large enough sample to use the class as a unit. Furthermore, the investigator did not use groups smaller than class size as instructional units because of the

interest in the reaction of class size groups to the treatments. Thus, the individual was used as the unit of analysis and the results will be interpreted in light of this.

Chapter V

EXECUTION OF THE STUDY

Introduction

The preceding chapter described the development of the study. This chapter reports the events which occurred from the time of the selection of the population through the time of the collection of the data. A description of the population, summaries of the teach-test procedure and the instructional treatments, the criteria and procedure for determining the aptitude levels, and the results from the observations are included.

Population

The study was conducted at Hartland Elementary School, Hartland, Wisconsin. Hartland has a population of about 3000 and is located in Waukesha County approximately 30 miles west of Milwaukee. Although primarily a rural community many of its citizens are employed in Milwaukee or other nearby towns.

Hartland Elementary School has 554 students enrolled in kindergarten through eighth grade. It has its own school board which functions independently of the other elementary schools in the county. The primary classes are in a new part of the building; the rooms are pleasant and well equipped.

Because of the size of the school all of the second and third grade students were used. This provided 110 subjects, 44 from the second grade and 66 from the third grade. Both grades were using Addison Wesley's Elementary School Mathematics, 1968, for mathematics. The second grade was also using AAAS's Science: A Process Approach, 1968, for the first time.

Prestudy Arrangements

Selection of teacher. Miss Marcia Dana, a certified elementary school teacher experienced in teaching primary grades, was selected to teach the instructional unit of the teach-test procedure and both treatments. At the time of the experiment she was a Project Specialist at the Wisconsin Research and Development Center. This position involved writing activities for DMP Levels I-IV (kindergarten - second grade) and working with teachers in developmental schools. Thus, she was familiar with DMP's approach and philosophy.

In developing the activities for the treatments and for the instructional unit for the teach-test procedure, Miss Dana gave valuable assistance to the investigator in matters which concerned classroom management or selection of suitable activities. Hence, other than a daily review of the purposes and procedures for the next day, it was unnecessary to give her any special training.

Initial contact with the school. The principal of Hartland Elementary School was contacted about the possibility of conducting the

study in his school. When the type of mathematics program which DMP purports and the type of study were explained, he readily consented. The proposed dates, November 22, 1971 through December 10, 1971 and one day in January, 1972, were acceptable.

Meeting with the teachers involved. Approximately one week before the beginning of the study, both the investigator and the experimental teacher met with the six second and third grade teachers and an assistant administrator of the system. At this time the study was explained to the teachers and an invitation to attend any of the classes was extended. Also, the investigator and the experimental teacher became familiar with the school and its practices. The teachers appeared to be most cooperative; an observation which was confirmed throughout the study.

At this meeting the necessary physical arrangements and schedules were made. It was decided that the teach-test procedure would be conducted in the individual classrooms for the three days, Monday through Wednesday, before the Thanksgiving holidays. A schedule of times for each class was arranged which caused the least possible interruption to the regular schedules of the teachers. All the posttesting was scheduled for Wednesday morning.

Two teachers volunteered their classrooms for the instructional treatments. Both rooms were approximately the same size and had about the same educational climate. However, one was a second grade classroom and the other was a third grade classroom. Because it was incon-

venient for the teachers to change the treatment groups daily, it was decided to change only at the end of the first week. Two forty-minute periods, one from 9:00 to 9:40 and the other from 10:35 to 11:15, were scheduled. For the first week Treatment U was to be held from 9:00 to 9:40 and Treatment N from 10:35 to 11:15. During the second week Treatment N was to be held first and Treatment U second. A summary of these arrangements is shown in Figure 5.1.

	Room 1 9:00 - 9:40	Room 2 10:35 - 11:15
1st Week	U	N
2nd Week	N	U

Figure 5.1. Physical and Temporal Arrangement for Treatments

The teachers were told that on the Monday morning after Thanksgiving they would be notified about which students were selected for each of the treatments. By working with the teachers at their grade level they agreed to make the necessary arrangement for the students who were not involved during the two periods. This involved combining classes when one of their level's rooms was in use and planning lessons that would not be crucial if missed by those participating in the experiment.

Summary of the Teach-test Procedure

The plan and purposes of the teach-test procedure were explained

in Chapter IV. The lesson plans, activity sheets and journal may be found in Appendix A and the observations O_1 and O_2 may be found in Appendix B. A summary of the three days is given here.

Monday, November 22. In each of the six classes the first twenty minutes were spent on the pretest, Observation O_1, and the last twenty minutes were spent on Lesson 1. There were 101 students present for these sessions. The day went as planned; there was no difficulty in administering the test and the children were responsive to the lesson.

Tuesday, November 23. Each class spent 40 minutes on Lesson 2 of the teach-test procedure. Both parts of the lesson kept the children's attention. It was evident from some individuals' responses that they were still centering on number and ignoring the unit. Likewise, there appeared to be many who were using the relationships between the given units in making comparisons.

Wednesday, November 24. Although 98 students were present for the posttest, Observation O_2, not all of these had been present for the previous two days. Complete data for 90 students were collected. Twenty minutes were allotted in each class for the test. The schedule was arranged so that all the tests could be administered in the morning. It was interesting to note that the second graders moved through the test more rapidly than the third graders did.

Results of O_1 and O_2

Observations O_1 and O_2 were used to determine the aptitude levels.

Before discussing this determination the results of these two observations are given and discussed here. These results are given for the ninety students who were present for all three days of the teach-test procedure. Their scores on O_1 and O_2 along with their grade, age and sex are reported in Table E.1 of Appendix E. The first two items on both forms O_1 and O_2 were used to test prerequisite behaviors and for warm-up items. Only the last twenty items, the items testing the objectives of the teach-test procedure, are considered in this analysis as well as in determining the levels of aptitude.

The means, standard deviations, Hoyt reliability coefficients and standard error for O_1 and O_2 are shown in Table 5.1.

Table 5.1

DESCRIPTIVE STATISTICS FOR O_1 AND O_2

Observation	Mean	Standard Deviation	Hoyt Reliability	Standard Error
O_1 (pretest)	9.244	2.079	.340	1.646
O_2 (posttest)	12.544	3.336	.703	1.772

As indicated in Table 5.1 the difference between the means of O_1 and O_2 was 3.3. A *t*-test was used to test the significance of this difference. A *t* value of 7.921 showed that this difference was significant at the $p < .001$ level.

The Hoyt reliability coefficient was determined by using an internal criterion, the score on the observation. The item analysis

for O_1 and O_2 may be found in Tables F.1 and F.2, respectively, of Appendix F. An examination of the item analysis for O_1 revealed that there were eight items with extremely low r biserials. The same eight items had slightly higher r biserials on O_2 but were the lowest on this observation also.

These eight items were the only items which a child would have answered correctly if he was always making the comparison based on number alone. These items also had low levels of difficulty. Thus, it appears that these items could account for the low reliabilities. (A separate analysis of these items and of the remaining twelve is reported and discussed in Chapter VII.) While there is danger in using unreliable instruments, in this case the investigator did not feel that this danger was crucial for two reasons. First, the low reliability was probably due to the eight items which measured a different ability from the other items. Second, the determination of the aptitude levels depended mainly on the score on O_2 whose reliability was much higher. This second point is discussed in more detail after the aptitude level determination is explained.

Determination of Aptitude Levels

The initial planning called for the three levels of aptitude to be determined by specifying a mastery level. Different combinations of mastery and non-mastery on the pretest and posttest were designated as the levels. It was evident from the pilots of the teach-test

procedure and from the teach-test portion of this study that there were three levels of achievement on each test. There was the mastery group, but there also was a group who had attained many of the behaviors without reaching mastery and a group who definitely was non-mastery. This middle group, the group which had not quite reached mastery, is what DMP classifies as "progressing toward mastery". It was decided that students in this category on the posttest or on the pretest who fell to non-mastery on the posttest would be dropped from the instructional treatments. Thus, the design of the study becomes similar to what Cronbach and Snow (1968, p. 21) call an extreme group design. In an extreme group design the middle group is eliminated. This decision assured that there would be more difference between the Levels I and II.

Furthermore, it was evident that the behavior of some students was changing from the pretest to the posttest. Since Level I was to consist of those students who were at a low level of attainment and who were not affected by the teach-test treatment, it was decided to add the restriction that those in Level I should have evidenced no real change. Likewise, since Level II consisted of those who changed during the teach-test treatment, the restriction that these subjects must have evidenced change was added.

Thus, in the final determination of the aptitude levels two criteria were used. First, the level of attainment of the specified behaviors and second, the amount of change evidenced from pretest to posttest. A mastery level of 75% was set. Any student in the 55-75% range was

classified as progressing toward mastery and any student at the 55% or below level was classified as a non-mastery student. A change of more than 15% was necessary to indicate a change due to instruction. Using these criteria the levels as defined are specified as follows:

Level III consists of those students who had attained mastery on both the pretest and the posttest. Thus, their score on both O_1 and O_2 must be 75% or better (15 points or better).

Level II consists of those students who had not attained mastery on the pretest, but had attained mastery on the posttest and had evidenced a change of more than 15%. Thus, their score on O_1 must be less than 75% (less than 15 points) and must be 75% or better (15 points or better) on the posttest O_2 and their score must have changed more than 15% (a change of more than 3 points).

Level I consists of those students who were at the non-mastery level on both tests and who evidenced no major change. Thus, their scores on both O_1 and O_2 must be 55% or less (11 points or less) and their score must have changed no more than 10% (2 points).

These levels are summarized in Figure 5.2.

	O_1	O_2	Change
Level III	$s(O_1) \geq 15$	$s(O_2) \geq 15$	(no restriction)
Level II	$s(O_1) < 15$	$s(O_2) \geq 15$	$s(O_2) - s(O_1) > 3$
Level I	$s(O_1) \leq 11$	$S(O_2) \leq 11$	$\lvert s(O_2) - s(O_1) \rvert < 3$

[Note: $s(O_1)$ means score in points on O_1, etc.]

Figure 5.2. Criteria for Aptitude Levels

These specifications eliminated any student who

(a) was at mastery on the pretest, but not on the posttest

(b) was progressing toward mastery on the pretest, attained, mastery on the posttest, but did not evidence enough change,

(c) was progressing toward mastery on both tests,

(d) was progressing toward mastery on the pretest, but was nonmastery on the posttest,

(e) was nonmastery on the pretest and only progressing toward mastery on the posttest, or

(f) was nonmastery on both tests, but evidenced too much change.

Figure 5.3 summarizes all the logical combinations of mastery (m), progressing toward mastery (p), and nonmastery (n) with the added restrictions of change (c). It also shows how each combination was classified for this study and the number of subjects in each classification.

As Figure 5.3 shows these classifications produced 2 students in Level III, 27 in Level II and 32 in Level I. Table E.2 in Appendix E reports the scores on O_1 and O_2 for those three levels. Twenty-nine students were eliminated for one of the six reasons (a-f) listed in the previous paragraph. Only one subject was eliminated for each of the reasons, c and d; two were eliminated for reasons a and b; fourteen for reason e and nine for reason f. Table E.3 in Appendix E gives their scores, as well as the reasons why they were eliminated.

O_1 level of mastery	O_2 level of mastery	Change	Classification	Number of Students
m	m		Level III	2
m	$\sim$ m		(a)	2
$\sim$ m	m	$c > 3$	Level II	27
$\sim$ m	m	$c < 3$	(b)	2
p	p		(c)	1
p	n		(d)	1
n	p		(e)	14
n	n	$\|c\| < 3$	Level I	32
n	n	$\|c\| > 3$	(f)	9

[Note: $\sim$ m means p or n.]

Figure 5.3. Possible Combinations of Mastery Levels on O_1 and O_2 and Classifications of These Combinations

These classifications were arbitrary with respect to the decisions for mastery level and amount of change. Once these decisions were made then the classification of an individual was determined. Due to errors of measurement two types of errors were possible. First, subjects may have been eliminated who should have been retained and second, subjects may have been retained who should have been eliminated. Because the remainder of the study depended upon those who were retained the second type of error was more crucial than the first. Although there is no assurance that this type of error was

not made, the method of selection should have held it to a minimum. The first criteria of selection depended on Observation O_2, the more reliable instrument. After a subject met this criteria he also had to meet the criteria for change. Considering the standard error for O_1 and O_2 (see Table 5.1) there is no evidence to support the statement that a subject in Level II who had to gain at least four points did so because of error in measurement. On the other hand the subjects in Level I were permitted a change of two points, a change that may have been due solely to error in measurement. Furthermore and more importantly, it is essential that those retained were in the correct levels. Since a score of 15 or more on O_2 was required for Level II subjects and a score of 11 or less on O_2 was required for Level I subjects, there is no reason to believe that a subject classified in one level should have been in the other.

Because there were only two students in Level III they were dropped from the instructional treatments. Thus, only two levels were retained. The means and standard deviations on O_1 and O_2 for these two levels are reported in Table 5.2.

Looking at Table 5.2 one observes that there is not much difference between the means of Level I and Level II on O_1. A t-test ($t = 1.993$) showed that the difference between these means was not significant at the $p < .01$ level. However, the difference between the means of Level I and Level II on O_2 was significant at the $p < .01$ level ($t = 24.754$). It appears that these two groups were not dif-

ferent in the abilities tested before the teach-test treatment, but differed significantly in these abilities afterwards.

Table 5.2

DESCRIPTIVE STATISTICS FOR LEVELS I AND II

	Number	Observation O_1		Observation O_2	
		Mean	Standard Deviation	Mean	Standard Deviation
Level I	32	8.594	.701	9.438	1.059
Level II	27	9.222	1.474	16.556	1.010

Treatments' Population

The students in Level I and Level II were randomly assigned to Treatments N and U. Treatment N had 16 Level I students and 13 Level II students and Treatment U had 16 Level I students and 14 Level II students. Although no hypotheses were proposed concerning grade or sex, the groups are further described here according to sex and grade. There were 9 boys and 20 girls in Treatment N and 11 boys and 19 girls in Treatment U. It is interesting to note that although there were fewer boys in the treatments there were more boys than girls in Level II. There were 9 second graders and 20 third graders in Treatment N and 10 second graders and 20 third graders in Treatment U. Most second graders were in Level I; however, as many third graders as second graders were in this level. Figure 5.4 is a complete description of the sample.

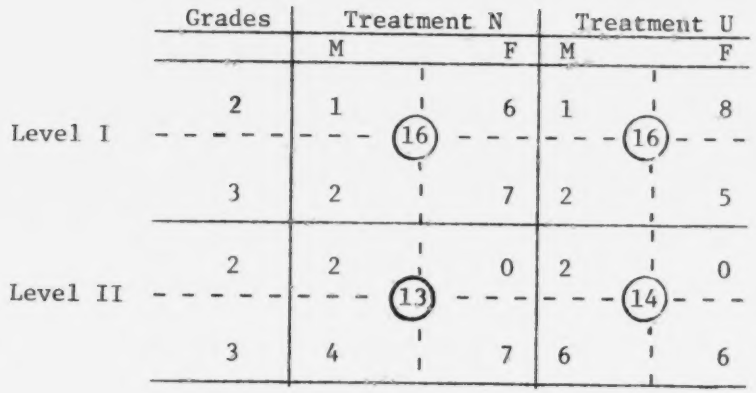


Figure 5.4. Description of Treatments' Sample According to Level, Age, Sex and Grade

Table E.5 in Appendix E contains the scores on O_1 and O_2 by treatment groups and levels.

Summary of the Treatments

The planned treatments were described in Chapter IV. A brief summary of the nine days of treatments is given here. The reader is referred to Appendix C for the lesson plans, the activity sheets and the daily journal for each treatment.

During the first week the subjects in Treatment U met from 9:00 to 9:40 and the subjects in Treatment N met from 10:35 to 11:15. For the second week the groups interchanged both times and rooms. The investigator felt that the time made little difference in the group's reactions, but that the third grade classroom was better equipped for thirty students.

There were few absentees; most of these were not for more than two days. Upon their return either the experimental teacher, the investigator or another student helped the absentee with the essentials of what he had missed. No one was dropped from the study because of absences during the treatment period.

The investigator was present in the classroom each day to observe the students and to make certain that the objectives of each lesson were adequately covered. If it was felt that an objective needed additional attention the remaining lessons were modified accordingly. This occurred only a few times and the adjustments made were minor. At times the type of activity was changed to provide a change of pace. For example, a game from the last activity was played on the seventh day by some of the students in Treatment U. The daily journal records such changes. It should be noted that more activities usually were planned for each day than the investigator felt were necessary. Thus, the last part of many lessons were not done on the day specified or at a later time.

There were no major interruptions. The teachers involved and the principal and his staff remained cooperative throughout the two weeks. The teachers often observed the classes when they did not have other responsibilities.

The children adjusted quickly to the new routine; they remained enthusiastic but reacted naturally to the whole experiment. In particular, they enjoyed working at stations which involved moving around

the room, contests which involved estimating and the variety of materials and pictures. There were no apparent differences between the students in the two treatment groups. If one day the investigator felt the lesson was more successful in Treatment N, then it was probable that the next day it would be just the opposite.

There were usually fewer essential points to be covered in Treatment N than in Treatment U so that class appeared more relaxed. However, there was always plenty to be done; the children were content and enjoyed it, but it was often not challenging enough for the better students. In Treatment U the better students were challenged--a challenge that was often over the head of a slower student. But the slower student was not frustrated because he did not even realize the challenge and could arrive at an answer which satisfied him.

The investigator was satisfied that both groups were presented with the same type of activities and the same amount of manipulative work. The questions asked and the discussions, not the instructional mode, accounted for the difference between the treatments.

Results of Observations O_3 - O_5

Three dependent measures were taken: an immediate achievement measure (Observation O_3), a transfer measure (Observation O_4) and a retention measure (Observation O_5). These instruments may be found in Appendix D. Both O_3 and O_4 were administered on the tenth day, December 10, 1971. Observation O_3 took approximately 40 minutes;

after an hour break Observation O_4 was given. There was little difficulty in administering O_3, the directions were clear and the students understood what was being asked. Observation O_4 was more difficult to administer due to the unfamiliarity of the content, to the more complicated directions and to the demand for a second period of concentration within one day. Three students were absent for O_3 and O_4. Observation O_5 was administered on January 12, 1972. It was the same instrument that had been used for O_3 and administering it presented no difficulty. The three students who were absent for O_3 and O_4 were present, but two others were absent. All five of these students were dropped from the final analysis of variance.

The raw data for these observations are given in Table E.5 in Appendix E. Because O_5 is made from the same instrument as O_3, the scores for O_5 follow those for O_3. Hence, for each student the achievement, retention and, then, transfer scores are given.

The descriptive statistics for $O_3 - O_5$ are given in Table 5.3.

Table 5.3

DESCRIPTIVE STATISTICS FOR $O_3 - O_5$

Observation	Number of Items	Mean	Standard Deviation	Hoyt Reliability	Standard Error
O_3 (achievement)	30	16.5179	6.8357	.8948	2.1797
O_4 (transfer)	25	6.8214	2.9487	.5432	1.9526
O_5 (retention)	30	17.7368	7.5298	.9215	2.0736

The mean score on O_5 was slightly higher than on O_3. This difference, as will be seen in Chapter VI, was due to the increase in Treatment U's scores from O_3 to O_5. The Hoyt reliability coefficients for O_3 and for O_5 were high. The mean score for O_4 was extremely low. The transfer test(O_4) proved to be unexpectedly difficult both from the standpoint of individuals and individual items. Only 3 subjects answered correctly over half of the items and the highest score was 17. Likewise, only three problems were correctly answered by more than half the subjects. The reliability of O_4 was low. Thus, because of the low mean and reliability of O_4 any interpretation of further statistical analysis arising from this observation must be made cautiously. The item analyses for observations O_3 - O_5 may be found in Tables F.3 - F.5, respectively, in Appendix F.

The statistical tests of the hypotheses of this study which involve the three observations O_3 - O_5 are reported in Chapter VI. Chapter VII contains the conclusions reached based on the statistical analyses.

Chapter VI

DATA ANALYSIS

Introduction

This chapter presents the data analyses used for testing the hypotheses of this study and an analysis of retention ratios. The research hypotheses which were stated in Chapter IV are summarized here. Each primary hypothesis stated that there is a significant interaction between aptitude and treatment. There were three such hypotheses; one for each dependent measure. The secondary hypotheses each stated that there is a significant main effect. There were six secondary hypotheses; a main effect due to aptitude hypothesis and a main effect due to treatment hypothesis for each of the three dependent measures.

Each of these hypotheses is discussed within the section dealing with the appropriate dependent measure. In each case it was planned to use a 3 X 2 ANOVA but since only two levels of aptitude were found each null hypothesis was tested by a 2 X 2 ANOVA. A significance level of .05 was set. After the discussion of the hypotheses an examination of the retention ratios is reported.

Achievement Measure

One dependent variable was achievement; this was measured by observation O_3. The instrument used for O_3 consisted of 30 items which tested the objectives of the treatments and was described in Chapter V. The

raw data for O_3 may be found in Table E.5 in Appendix E. The maximum possible score on O_3 is 30; the scores for this population ranged from 5 to 28 with a mean of 16.5. The descriptive statistics for O_3 related to aptitude levels and treatments are reported in Table 6.1.

Table 6.1

DESCRIPTIVE STATISTICS FOR O_3: APTITUDE, TREATMENT, AND APTITUDE BY TREATMENT GROUPS

Group	Number of Subjects	Mean	Standard Deviation
Aptitude			
I	29	14.069	6.216
II	25	19.320	6.517
Treatment			
U	28	20.786	5.412
N	26	11.884	4.950
Aptitude by Treatment			
I U	16	18.125	5.328
I N	13	9.077	2.397
II U	12	24.333	3.025
II N	13	14.692	5.313

Figures 6.1 and 6.2 give a graphic picture of the mean scores of achievement for the aptitude by treatment groups. Figure 6.1 shows that Level II subjects' scores were higher than Level I subjects' scores in each of the treatment groups. In Figure 6.2 it appears that Treatment U was more effective than Treatment N for both levels of aptitude. However, note that in each graph the differences between the corresponding ordinates are approximately equal. It appears that there is little chance of any interaction.

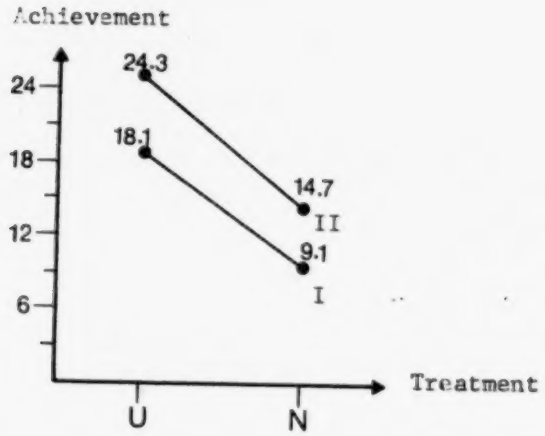


Figure 6.1. Aptitude Effect on Achievement

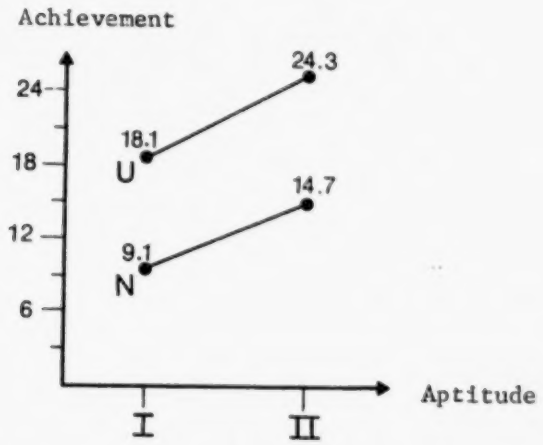


Figure 6.2. Treatment Effect on Achievement

These observations were examined by testing the following null hypotheses:

H.1a: The difference between the mean score on achievement of the Treatment U Level I group and the mean score on achievement of the Treatment U Level II group is equal

to the difference between the mean score on achievement of the Treatment N Level I group and the mean score on achievement of the Treatment N Level II group. (In other words, there exists no interaction between the aptitude level and the treatment as measured by achievement.)

H.1b: The mean score on achievement for Level I equals the mean score on achievement for Level II.

H.1c: The mean score on achievement for Treatment U equals the mean score on achievement for Treatment N.

The first hypothesis H.1a corresponds to the primary research hypothesis for achievement and the other two correspond to the two secondary hypotheses for achievement. Statistics relevant to testing each of these null hypotheses are presented in Table 6.2.

Table 6.2

ANALYSIS OF VARIANCE FOR O_3 (ACHIEVEMENT)

Source	df	MS	F	p <
Aptitude X Treatment	1	1.732	.063	.8032
Aptitude	1	468.086	25.055	.0001
Treatment	1	1166.020	62.414	.0001
Error	50	18.682		

As can be seen from Table 6.2 hypothesis H.1a can not be rejected. Statistically, the interaction as measured by achievement between aptitude and treatment was not significant. Both hypotheses H.1b and H.1c are rejected. There is statistical evidence that both the aptitude effect and the treatment effect were significant. Thus, the analysis of variance supports the observations made from Figures 6.1

and 6.2.

Transfer Measure

The second dependent variable was transfer; this was measured by observation O_4. The instrument used for O_4 was discussed in Chapter V. The raw data for O_4 may be found in Table E.5 in Appendix E. Out of a possible score of 25 the highest score was 17 and the lowest score was 0. From the mean of 6.8 one sees that the test was extremely difficult for this population. The descriptive statistics for O_4 related to levels and treatments are reported in Table 6.3.

Table 6.3

DESCRIPTIVE STATISTICS FOR O_4: APTITUDE, TREATMENT AND APTITUDE BY TREATMENT GROUPS

Group	Number of Students	Mean	Standard Deviation
Aptitude			
I	29	5.690	2.054
II	25	8.000	3.277
Treatment			
U	28	7.25	7.009
N	26	6.23	3.081
Aptitude by Treatment			
I U	16	6.688	1.852
I N	13	4.462	1.613
II U	12	8.000	3.384
II N	13	8.000	3.364

The graphs of the mean scores on transfer for the aptitude by treatment groups are found in Figures 6.3 and 6.4.

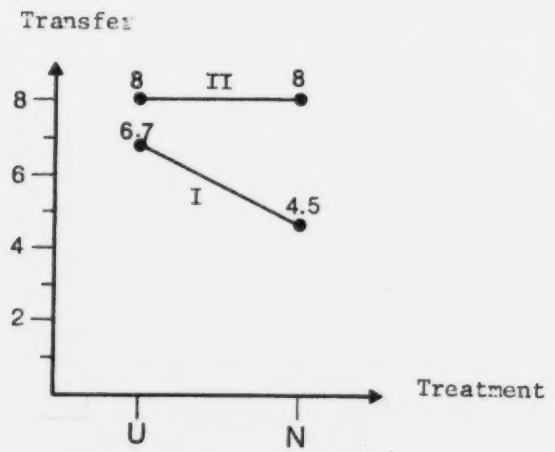


Figure 6.3. Aptitude Effect on Transfer

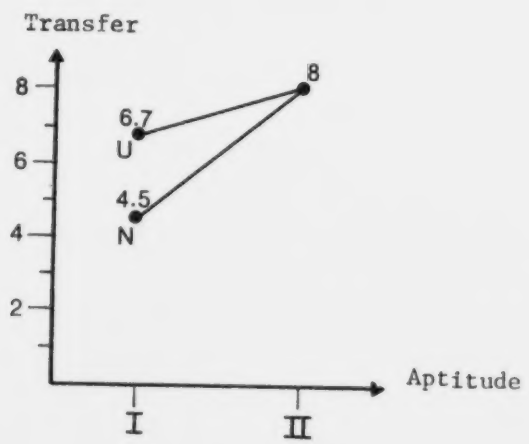


Figure 6.4. Treatment Effect on Transfer

From Figure 6.3 it appears that Level II was superior to Level I for each of the treatments. Thus, there appears to be a significant main effect due to aptitude. Figure 6.4 shows that Treatment U was slightly superior to Treatment N but only for Level I students.

These differences were statistically examined when the following null hypotheses were tested:

H.2a: The difference between the mean score on transfer of the Treatment U Level I group and the mean score on transfer of the Treatment U Level II group is equal to the difference between the mean score on transfer of the Treatment N Level I group and the mean score on transfer of the Treatment N Level II group.

H.2b: The mean score on transfer for Level I equals the mean score on transfer for Level II.

H.2c: The mean score on transfer for Treatment U equals the mean score on transfer for Treatment N.

Hypothesis H.2a corresponds to the primary research hypothesis for transfer and H.2b and H.2c correspond to the secondary ones.

Table 6.4 shows the results of the analysis of variance used to test these hypotheses.

Table 6.4

ANALYSIS OF VARIANCE FOR O_4 (TRANSFER)

Source	df	MS	F	p <
Aptitude X Treatment	1	16.534	2.344	.1321
Aptitude	1	76.663	10.869	.0019
Treatment	1	19.005	2.694	.1070
Error	50	7.053		

From the analysis reported in Table 6.4 the following results were found. The F ratio used to test hypothesis H.2a was 2.344 which was not significant at the .05 level. Thus, hypothesis H.2a is not

rejected. There is statistical evidence of a significant aptitude main effect; however, the treatment main effect hypothesis H.2c is not rejected.

Retention Measure

The third dependent variable was retention; this was measured by observation O_5. The instrument used for O_5 was the same as the one used for O_3 and is described in Chapter V. The raw data for O_5 is in Table E.5 in Appendix E. The scores ranged from 5 to 30 and the mean was 17.7. The descriptive statistics associated with O_5 for aptitude, treatment and aptitude by treatment groups is presented in Table 6.5.

Table 6.5

DESCRIPTIVE STATISTICS FOR O_5: APTITUDE, TREATMENT AND APTITUDE BY TREATMENT GROUPS

Group	Number of Subjects	Mean	Standard Deviation
Aptitude			
I	29	15.310	7.087
II	25	20.560	6.777
Treatment			
U	28	22.964	4.834
N	26	12.115	5.075
Aptitude by Treatment			
I U	16	20.625	4.440
I N	13	8.769	2.920
II U	12	26.083	3.450
II N	13	15.462	4.612

Patterns of differences similar to those for the achievement

measure can be seen in Figures 6.5 and 6.6. It appears that there are strong aptitude and treatment effects, but no interaction.

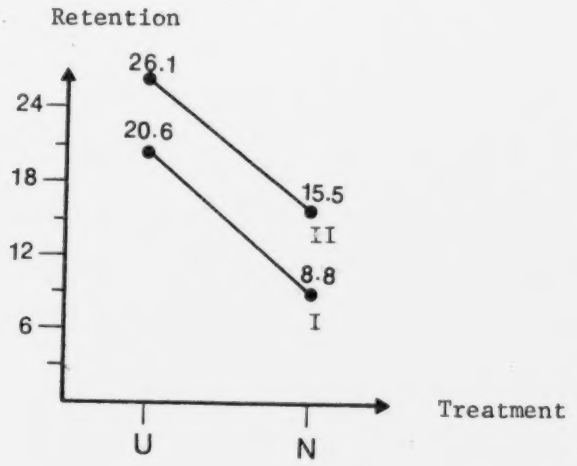


Figure 6.5. Aptitude Effect on Retention

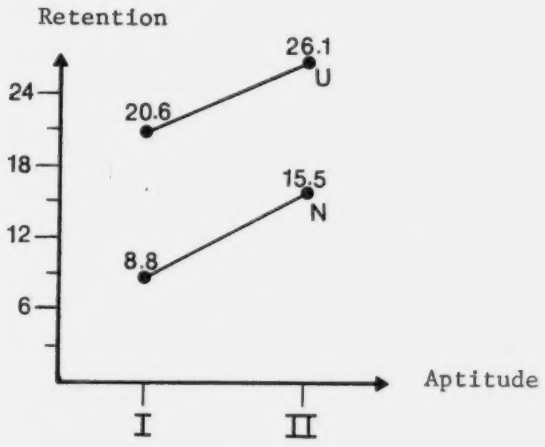


Figure 6.6. Treatment Effect on Retention

These observations were examined statistically when the following null hypotheses were tested:

H.3a: The difference between the mean score on retention of the Treatment U Level I group and the mean score on retention of the Treatment U Level II group is equal to the difference between the mean score on retention of the Treatment N Level I group and the mean score on retention of the Treatment N Level II group.

H.3b: The mean score on retention for Level I equals the mean score on retention for Level II.

H.3c: The mean score on retention for Treatment U equals the mean score on retention for Treatment N.

Again, the first hypothesis is the one of primary interest since it corresponds to the research interaction hypothesis for retention. The other two correspond to the main effects research hypotheses for retention.

The statistics associated with the analysis of variance used to test hypotheses H.3a, H.3b, and H.3c are reported in Table 6.6.

Table 6.6

ANALYSIS OF VARIANCE FOR O_5 (RETENTION)

Source	df	MS	F	p <
Aptitude X Treatment	1	5.081	.324	.5718
Aptitude	1	490.331	31.263	.0001
Treatment	1	1707.081	108.841	.0001
Error	50	15.684		

As Table 6.6 shows, the primary null hypothesis H.3a is not rejected. Again there is no evidence of any interaction. Both of the

hypotheses H.3b and H.3c are rejected. There is statistical evidence of a main effect due to aptitude and a main effect due to retention. Thus, the analysis of variance supports the observations made from Figures 6.5 and 6.6.

Retention Ratios

In addition to the hypotheses tested in this study, questions about retention were asked. In particular, the following two questions were posed:

1. To what extent are the performances on achievement correlated with performances on retention measured four and one-half weeks later?

2. How much retention was there?

For the entire group of 54 subjects the correlation between achievement and retention was .72. Data for individuals on the achievement observation O_3 and the retention observation O_5 are reported in Table 6.7. From the achievement measure to the retention measure 35 subjects' scores improved, 14 scores declined, and 6 scores remained the same. Of the 34 scores that improved, the range of improvement was from 1 to 8 raw score points. Of the 14 scores which declined, the range was from 1 to 6 raw score points. Retention ratios (amount retained divided by amount achieved) varied from .60 to 2.00 with 40 subjects having retention ratios of 1.00 or better.

The amount retained for the entire group is indicated by the estimated mean scores. The mean score on O_3 was 16.5 and the mean score on O_5 was 17.7. The retention ratio of 1.07 indicates that the subjects as a group did better on the retention observation than on

the achievement measure.

Table 6.7

INDIVIDUAL RESULTS ON ACHIEVEMENT (O_3), RETENTION (O_5) AND RETENTION RATIOS

	Treatment U			Treatment N		
	O_3	O_5	Retention Ratio	O_3	O_5	Retention Ratio
LEVEL I	13	20	1.54	5	8	1.60
	23	20	.87	7	9	1.29
	16	19	1.19	11	7	.64
	5	10	2.00	15	13	.87
	16	23	1.44	9	7	.77
	18	20	1.11	10	15	1.50
	26	26	1.00	10	6	.60
	15	20	1.33	9	11	1.22
	19	27	1.42	10	7	.70
	21	24	1.14	7	7	1.00
	14	16	1.14	8	8	1.00
	23	16	.70	8	5	.63
	25	24	.96	9	11	1.22
	19	25	1.32			
	15	17	1.13			
LEVEL II	24	18	.75	13	14	1.08
	25	25	1.00	11	10	.91
	22	27	1.23	9	11	1.22
	28	30	1.07	24	21	.88
	25	26	1.04	15	15	1.00
	20	25	1.25	20	25	1.11
	28	29	1.03	17	20	1.18
	26	28	1.08	25	23	.92
	26	29	1.11	9	12	1.33
	24	25	1.04	11	10	.91
	26	29	1.11	12	15	1.25
	18	22	1.22	13	15	1.15
				12	12	1.00

Summary

Each of the null hypotheses which corresponded to a primary research hypothesis concerning interaction is not rejected. That is, no

significant interaction was found for any of the three dependent measures. All except one of the null hypotheses which corresponded to the secondary research hypotheses are rejected. That is, there were significant main effects, both aptitude and treatment, for each of the dependent measures with the exception of the treatment main effect for transfer.

The examination of retention ratios revealed little difference in retention due to treatment or aptitude. The extremely high retention ratios indicated not only that most students retained what they had learned but also that their performance improved.

Chapter VII

CONCLUSION TO THE THESIS

Introduction

After giving a brief summary of the study and its limitations, this chapter discusses the conclusions, the implications for curriculum development and the recommendations for future research.

Summary

The main purpose of this study was to examine the interaction of two treatments on measuring area with various levels of aptitude. The research strategy adopted was that of an aptitude-treatment interaction study. Most of the past educational ATI studies fit what Salomon (1971) described as the preferential model. This model prescribes treatment differing on form or mode and capitalizes on the learners' best general capabilities. This study approached the ATI question in a different manner. General capabilities were not used as measures of aptitude and the treatments did not differ in form or mode. Aptitude was defined in terms of the individual's ability to learn specific concepts associated with a unit of length measurement. The treatments were designed to differ only in their emphasis on the unit of area measurement. More specifically, the question asked was: In what manner does the ability of children to learn concepts associated with a unit of length affect the

extent to which they attain concepts associated with area and a unit of area for each of the two given treatments?

In order to determine this ability 90 second and third graders were subjected to a teach-test procedure. This procedure consisted of a pretest, a brief instructional treatment and a posttest all of which tested or taught unit of length concepts. The results of the two tests were used to determine the aptitude levels. Although three levels of aptitude were expected only two subjects met the criteria for the highest level. They were dropped from the remainder of the study along with those students who did not fit the definition of either of the other two levels. Twenty-seven and thirty-two students were classified as Level II and Level I, respectively. The students in each of these levels were randomly assigned to one of two treatments, Treatment U or Treatment N.

Both treatments had the same behavioral objectives, the same teacher, the same duration (9 days) and the same mode of instruction. They ffered in the amount of emphasis on the unit of measure for area; Treatment U emphasized the unit and Treatment N did not. After the treatments three measures, achievement, transfer and retention, were taken. These measures were used to test hypotheses about the interaction of the aptitude levels and the treatments and about the main effects of aptitude and of treatments.

No significant interactions were found between the levels and treatments on any of the measures. There were significant main effects due to level of aptitude and to treatment for achievement and retention measures. The only significant main effect for the transfer measure

was aptitude.

Limitations

While the sources of internal validity (Campbell and Stanley, 1963) were controlled through the design of the study, some of the sources of external validity which permit generalization were not amenable to control. The first source of external validity which was not controlled was the interaction of selection and treatment. The school selected was a rural school whose staff was most cooperative. Thus the results must be interpreted for this population. A second source of invalidity may have been what Campbell and Stanley (1963) call reactive arrangements - "the patent artificiality of the experimental setting" (p. 20). While every effort was made to have a normal setting there is no way to measure the effect of the experiment itself. In this case it was not felt that the students in the two groups reacted differently to the experimental setting. It was more probable that the biases of the investigator and of the teacher, or the unusual mathematical expertise of the teacher would have accounted for any reactive arrangement invalidity.

In interpreting the results one must also consider the reliability of the instruments. These reliabilities were reported in Chapter V. The reliabilities of the achievement and retention observation were respectable, but the reliability of the transfer test as well as the level of the difficulty of its items make any transfer findings suspect.

Finally, the statistical analysis calls for independent observations. The use of individuals for the unit of analysis makes any results questionable. However, the clear cut results of treatment and

aptitude level effects reduces the questionability. This limitation was discussed more fully in the statistical analysis section of Chapter IV.

Conclusions

This section discusses the conclusions relevant to the teach-test procedure and then the conclusions relevant to the hypotheses of the study.

The teach-test procedure produced two distinct levels of aptitude. The mean scores for Level I and for Level II on the pretest, O_1, were 8.594 and 9.222, respectively. On the posttest, O_2, the mean score for Level I had not changed significantly ($p < .05$). However, the mean score for Level II was 16.556 which was a significant change ($p < .05$). Likewise, the difference between the posttest mean scores for the two groups was significant ($p < .05$). Furthermore, for the dependent measures of achievement and retention there was a significant aptitude effect. Thus, the teach-test method was successful in predicting the ability of students to learn concepts associated with measuring area.

These conclusions should be interpreted in light of the reliability coefficients .340 and .703 for the two observations O_1 and O_2, respectively. Neither reliability coefficient was extremely high. This problem first was addressed in Chapter V. A closer examination of the reliabilities was made in the following manner.

In both the instruments used for O_1 and O_2 there existed what appeared to be two categories of questions which could be answered correctly for different reasons. The first category consists of eight

questions which would be answered correctly if the child was only responding to the numbers given and ignoring the unit. For example, he was asked to order two objects which measured 8 and 6 of unit a, respectively. The second category consists of twelve questions which cannot be answered correctly if the child was only responding to the number and ignoring the unit. For example, he was asked to order two objects which measured 10 of unit a and 10 of unit b, respectively, when the two units in question were not equal.

Separate Hoyt reliability coefficients for each of these categories for each observation O_1 and O_2 were found. For both subscales of O_1 the reliability coefficients were .82 compared with the reliability coefficient of .34 for the entire scale. For O_2 the reliability of the first category of items (subscale 1) was .73 and the reliability of the second category of items (subscale 2) was .84 compared with the reliability of .70 for the entire scale. The item analyses for these separate subscales are found in Tables F.6 and F.7 in Appendix F.

It is important that decisions made from instruments are made from reliable ones. The subscales were more reliable than the entire scales for both O_1 and O_2. It was thus decided to compare the selection of aptitude levels based on the entire scale with the selection based on a subscale.

Since the items on the second category were more discriminating the scores of students on this category for both O_1 and O_2 were considered. By subtracting the 8 possible points which may have been attained on the first category items, new criteria for levels and for elimination are shown in Figure 7.1.

Using these new criteria only two students in Level II would not have been reclassified as Level II. Both of these would have been classified in Level III under this system. On the other hand 9 of the 32 Level I students would have been reclassified, only 3 of these would have been placed at a higher level and the other 6 would have been eliminated. Only five of the students who were originally eliminated would have been retained, two in Level II and three in Level I.

O_1	O_2	Change	Classification
$S(O_1)^* \geq 7$	$S(O_2) \geq 7$		Level III
$S(O_1) \geq 7$	$S(O_2) < 7$		(a)
$S(O_1) < 7$	$S(O_2) \geq 7$	$c > 3$	Level II
$S(O_1) < 7$	$S(O_2) \geq 7$	$c \leq 3$	(b)
$3 < S(O_1) < 7$	$3 < S(O_2) < 7$		(c)
$3 < S(O_1) < 7$	$S(O_2) \leq 3$		(d)
$S(O_1) \leq 3$	$3 \leq S(O_2) < 7$		(e)
$S(O_1) \leq 3$	$S(O_2) \leq 3$	$\|c\| < 3$	Level I
$S(O_1) \leq 3$	$S(O_2) \leq 3$	$\|c\| \geq 3$	(f)

*$S(O_1)$ indicates score on subscale 2 of O_1 and c indicates amount of change between O_1 and O_2 on subscale 2.

Figure 7.1. Classification for Aptitude Levels Based on Subscale 2 of O_1 and O_2

Figure 7.2 shows the mean scores on achievement for those students who were in Treatments U and N for both the new classification and the classification used in the study.

Classification Using Entire Scale			Reclassification Using Subscale 2		
	U	N		U	N
I	18.2	9.0	I	18.1	9.1
II	24.6	14.5	II	24.3	14.7

Figure 7.2. Mean Scores on Achievement According to Two Classifications

Since such small differences existed no further analysis was made on the achievement data. Because the retention data was so similar to the achievement and the transfer measure itself was not very reliable these were not examined in light of the reclassification.

There is no doubt that this subscale was more reliable than the entire scale. However, the conclusions reached using the entire scale for classification appear to be no different from the classification on the subscale. Furthermore, the subscale did not test all the behaviors desired. It was desirable, as was the case of the entire scale, that some items could be answered correctly by comparing only the numbers involved. Otherwise there was no way to determine the students who focused on the length of the unit alone and ignored the measure. The three students in Level I who would have been classified in higher levels on the subscale classification were apparently looking only at the length of the unit and not at the number of units, an error as crucial to

detect as the one of centering on measure alone. Thus, while it is desirable to use reliable instruments the investigator feels confident that the lack of high reliability in this case did not affect the final conclusions.

The analysis of the data related to the three primary hypotheses was reported in Chapter VI. All three analysis of variance tests revealed that there was no significant interaction between treatment and aptitude. Thus, the study failed to produce the desired interaction. However, this conclusion must be considered in regard to the fact that only two levels of aptitude were established by the teach-test procedure. The third level, the highest level of aptitude, was not present in this population. It was expected that this level of student would do equally well under either treatment which would have helped to produce the desired interaction.

The analysis associated with either the dependent measure, achievement, or the dependent measure, retention, substantiates the same conclusions. In examining the main effects hypotheses, it was found that Treatment U was significantly better than Treatment N for both levels of aptitude. In regards to achievement, Treatment U was so strong that there was little chance of interaction. The mean score for Level I students in Treatment U was higher than the mean score for Level II students in Treatment N. Thus, for either level, as far as achievement of the specified objectives, Treatment U is preferable to Treatment N. Likewise, there was a significant aptitude main effect; those in Level II did better than those in Level I. Retention data and associated analysis supports these conclusions.

The achievement retention data were examined more closely in relation to these three questions:

1. Is the amount of retention a function of the aptitude level?
2. Is the amount of retention a function of the particular treatment?
3. Is the amount of retention a function of the interaction of aptitude and treatment?

Table 7.1 includes the groups' mean scores on achievement (O_3) and retention (O_5) and the associated retention ratios.

Table 7.1

GROUP MEANS ON ACHIEVEMENT (O_3) AND RETENTION (O_5) AND RETENTION RATIOS

Group	O_3	O_5	Retention Ratios
Aptitude			
I	14.1	15.3	1.09
II	19.3	20.6	1.06
Treatment			
U	20.8	23.0	1.11
N	11.9	12.1	1.02
Aptitude by Treatment			
I U	18.1	20.6	1.13
I N	9.1	8.8	.97
II U	24.3	26.1	1.07
II N	14.7	15.5	1.05

As one can see from Table 7.1, for Level I there was an increase of .8 from achievement to retention and for Level II an increase of 1.3. The retention ratio for Level I was 1.09 and for Level II was 1.06 indicating that both groups' percentage of increase was approximately

the same. Looking at the individual data in Table 6.7 for these two groups one notes that nine students in Level I decreased in performance from O_3 to O_5 and five in Level II decreased. Thus, it appears that the aptitude level had little effect on the amount of retention.

The amount of retention for each treatment group is reported in Table 7.1. The mean score of Treatment U group increased by 2.2 from achievement to retention while Treatment N group's mean score only increased .2. The retention ratio for group U was 1.11 and for group N was 1.02. The individual data reported in Table 6.7 indicates a similar pattern. Only 4 subjects' scores decreased in Treatment U while 10 scores were lower on retention than on achievement in Treatment N. These retention ratios, the raw mean score gain and individual data seem to indicate that Treatment U was slightly more effective than Treatment N. This must be interpreted in light of the findings reported in the next paragraph.

If the amount of retention for each aptitude by treatment group is examined (see Table 7.1) one finds that the only group which decreased from achievement to retention was Treatment N Level I. On the other hand Treatment U Level I group showed the most increase. These groups appeared to contribute substantially to the conclusions reached in the previous paragraph about the effect of the treatment. While no further analysis was carried out, it appears that the amount of retention was affected to some extent by the interaction of aptitude and treatment.

The retention was extremely high for all groups. As far as the investigator could ascertain there was no teaching of the concepts

between the achievement observation and the retention observation. Much of the time between the two observations was a vacation period. There are several possible explanations for the high retention. It is possible that the achievement scores were lower than they would have been if the observation had not occurred prior to a major vacation. It is possible that the achievement test was a learning experience and this interacted with the retention observation. Or it is possible that subjects in Treatment U found the learning meaningful and subjects in Treatment N overlearned the objectives they learned; both of these are often used to explain high retention. The design of this study did not permit investigations of any of these possibilities.

The conclusions related to the transfer measure are not as clear-cut. Although there was a significant aptitude main effect, there was not a significant treatment main effect. The low reliability of the instrument and the high level of difficulty made interpretation of the results unsubstantiable.

Examining the instrument used for O_4 (transfer) more closely reveals that many of the items differed from the objectives in two dimensions; attribute and type of comparison question. The comparison question which asks about the inverse relation between an object and its measure or between an object and the unit of measure had not been asked previously. The instrument should be reconstructed to include more items that differ in only one dimension.

There were four items (4-7) on area which are very similar to the ones on the achievement test; the only difference is that units not called for in the treatments are used. The percent of correct responses

for the different groups of this study is shown in Figure 7.3.

	U	N
I	.59	.21
II	.64	.39

Figure 7.3. Percent of Correct Responses on Items 4-7 of Observation O_4

No statistical tests were run on this data; however, one notices that the ability to work with unfamiliar units favored Treatment U. It appears that the results of the transfer measure might have been different if more items had been of a near transfer type rather than of a far transfer type. That is, more items are needed which differ from the objectives on only one dimension rather than on two dimensions.

Implications for Curriculum Development

One of the purposes of this study was of a formative nature - to generate knowledge about mathematics instruction which could be used in a mathematics program. This is the purpose addressed here as information gained through subjective observations and objective testing is reported. All recommendations made here are in reference to the sample in this study. The background of the population involved should be carefully considered before adopting any of these recommendations.

First, the feasibility of teaching area concepts to second and third graders was examined. From observations during the treatments it is clear that many of the behaviors prerequisite to terminal

behaviors were easily obtained by the majority of the students. From the results of the achievement test it is evident that many second and third graders were successful in exhibiting the terminal behaviors. However, the significant main effect for aptitude shows that the ability of children at this age to attain these behaviors differs greatly. Although no statistical analysis was done on the difference in achievement due to grade the data shows some interesting trends. The mean score on achievement for the second graders was 13.2 and for third graders was 18.7. This is misleading unless one recalls that most of the second graders' were in Level I. Looking more carefully at the grades within levels, one notices on Level I that the second graders' mean is only one point lower than the third graders. Furthermore, the second graders in Treatment N did slightly better than the third graders. There were only three second graders in Level II; their scores were in the range of the third graders' scores. Second graders in Treatment U did slightly better than the third graders in Treatment N. Thus, it appears that the proper placement of these area concepts depends more on the child's ability as defined by the teach-test procedure of this study than on grade level and on the type of treatment. In designing a curriculum it is recommended that Treatment U be begun in the second grade and extended through the third grade.

When the achievement data was examined in terms of sex differences it was found that boys did better than girls in Treatment U, but the opposite was true for Treatment N. The direction of the difference in Treatment U could be expected since there was a much greater percentage

of boys in Level II than in Level I and a greater percentage of the girls were in Level I than Level II. Approximately the same division of boys and girls occurred in Treatment N, so the fact that girls did better than the boys is surprising. Again, no further statistical analysis was done; this information is reported only to support the recommendation that no difference in treatments seems to be necessary based only on sex.

Carpenter (1971) in discussing the implications from his study for instruction concludes:

> "These results do not imply that experiences with different units of measure should not be included in measurement topics. They do imply, however, that many young children will not master all the implications of different units by concentrating on measurement processes. If one is really concerned with mastery of measurement concepts with different units of measure, it would seem necessary to provide a wide range of experiences that help a child to focus on more than one immediate dimension. (p. 106)

This study verifies this observation made by Carpenter. The children in Treatment U who were given experiences with different units of measure behaved differently from those in Treatment N. The scores on the achievement and retention observations were significantly different for the two treatments.

In looking more carefully at these observations several other patterns of answers became obvious. One of the most striking was the recording of the unit when assigning measurements to areas. In both treatments whenever the teacher wrote the area she recorded both the number and the unit. One treatment did not overtly stress the writing

of the unit more than the other. However, the children in Treatment U saw the necessity for recording the unit. Almost without exception every child in Treatment U recorded the unit on every problem on both the achievement and the retention test. No students in Treatment N recorded the unit. Thus, if curriculum developers want the child to respond with the unit it seems to be necessary to present them with problems in which the unit makes a difference. In only presenting problems of comparing two areas which are covered with the same unit the child has no reason to focus on the unit and therefore does not record it.

There were eighteen problems on the achievement measure which asked the child to compare two regions which had been covered. These problems differed on two dimensions: (1) whether or not the regions were covered exactly and (2) whether or not congruent units were used to cover both regions. Figure 7.4 shows the number of problems in each category. Figure 7.5 shows the average level of difficulty (the average percentage of correct responses for each treatment and each level).

	Exact	Non-exact
Congruent	2	4
Non-congruent	8	4

Figure 7.4. Number of each type of comparison problems which involved two covered regions on O_3

		Exact		Non-exact	
		Treatment		Treatment	
		U	N	U	N
Congruent	I	88	91	56	29
	II	100	100	71	31
Non-congruent	I	70	21	36	10
	II	91	39	70	27

Figure 7.5. Precentage of correct responses to comparison problems which involved two covered regions on O_3

One notices in Table 7.5 that problems in which the coverings were exact and congruent units were used were the easiest for all groups. This is not surprising since only the numbers needed to be compared without regard to the unit. Likewise, the problems in which the coverings were non-exact and non-congruent units were used proved to be the most difficult for all groups. However, for Level II, Treatment U there was little difference in the difficulty due to the type of unit. A similar analysis of these scores on the retention measure indicated no difference between those covered with congruent and non-congruent units for this group. One may also observe that for Level I, Treatment N the second least difficult type of problem was the non-exact congruent. For every other group the exact, non-congruent was the second least difficult. This is consistent with the children's opportunity to learn and the results of the teach-test procedure. The subjects in Treatment

N had not been exposed to problems involving non-congruent units of area but had been given non-exact coverings. Thus, one would expect those in Level I, Treatment N to do better on the non-exact, congruent than on the exact, non-congruent problems. On the other hand the students in Level II, Treatment N were the ones who could handle non-congruent units of length in the teach-test procedure so it is not surprising that they could handle these problems with area.

Looking at the difference between the two treatment groups U and N and their difficulty levels on exact and non-exact problems one finds that 81% of the exact and 56% of the non-exact problem responses were correct for Treatment U students and 43% of the exact and 24% of the non-exact problem responses were correct for Treatment N. Thus, for both groups the non-exact problems were more difficult and the difference between the difficulty level was approximately the same.

However, if one makes a similar analysis between the congruent and non-congruent problems one finds that the level of difficulty between these two types was about the same for Treatment U. Seventy-three per cent of their responses were correct for the congruent type and 70 per cent for the non-congruent. But for Treatment N the non-congruent type was twice as difficult. Only 26 per cent of their responses were correct for the non-congruent type and 52 per cent were correct for the congruent.

In developing curriculum materials one should keep these results in mind. It appears that second and third graders are capable of handling problems involving non-congruent units, but they must be presented the opportunity. Also, special care must be given to problems

involving non-exact coverings since they appear to be more difficult than the exact ones.

Several subjective observations were made during the treatments concerning the unit. The children in Treatment U were more challenged by the problems they were presented. The better students in Treatment N had few experiences which stretched their ability. They probably could have learned everything presented in a much shorter time.

Further evidence of the ability of a child of this age to work with different units of measure was given by the teach-test procedure. As was shown by the results of O_1 and O_2 some of the children were and some were not capable after two days of instruction to handle the relationships between different units of measure. It is interesting to note that many of those students in Level I who could not work successfully with different units of length after the two day instruction period could handle such relationships with units of area after nine days of instruction and experience.

The results of the retention observation further supports the recommendation to design a curriculum which approaches area concepts as Treatment U did rather than as Treatment N. Furthermore, if one looks at the retention data in relation to the achievement data (see Table E.5 in Appendix E and Table 6.5 in Chapter 6) one finds more evidence to support this recommendation. The mean of the achievement observation was 2.2 lower than the mean of the retention observation for Treatment U, but only .2 lower for Treatment N. In Treatment U 22 of the 28 students showed an increase from the immediate achievement to the

retention while 12 of the 26 students in Treatment N increased. Nine Treatment U subjects increased more than three points (from four to eight points) but only one Treatment N subject increased more than three points. One of the aims of curriculum development is to construct programs which increase retention and it appears that Treatment U is slightly superior to Treatment N in regards to this.

Because of the difficulty level and the low reliability of the transfer test any recommendations with regard to transfer would be questionable. Since the third level of aptitude was not found in the teach-test procedure, there was no way to measure the transfer from length to area. It had been hypothesized that students in Level III would do equally well under either treatment because of their ability to transfer their knowledge about the unit of length to the unit of area. Thus, no specific recommendations regarding transfer can be made from objective observations.

Finally, some subjective observations about the type of instruction are relevant to curriculum development. The activities developed for this study were appropriate for this age group and were manageable. Many of the activities required extensive preparation, but much of the preparation could be done by students or the activities modified to simplify the preparation. The activities which held the children's interest longest were those which involved the children with materials and problems. The comparison problems were more motivational than ones which required merely assigning measurements. The children enjoyed the contests and games which added needed variety. The characters, the strange houses and the short stories told by the teacher to introduce

them served two purposes. First, they gave a way to introduce many ideas and to tie the activities together. Second, they held the children's interest and sparked their imaginations. When asked what they liked best the most responded that they liked the characters. Others liked the stations, the puzzles and the "snake" test (the teach-test procedure). Thus recommendations for curriculum development include providing variety through puzzles, games, contests, stories and story characters, providing activities which involve the children measuring objects in the room, providing problems to be solved and providing time for the children to discuss what they have observed.

Recommendations for Future Research

1) The final teach-test procedure did not produce the third level of aptitude. It might be worthwhile to replicate this procedure with another sample. There are three recommendations for changing the sample:

a) Select from third and fourth graders instead of from second and third graders.

b) Select from a DMP population; that is, select from a population which is familiar with a measurement approach and the type of activities presented in the treatments.

c) Select a larger sample of second and third graders from a different environment.

2) The retention data showed an increase from the achievement test. A future study should look at a retention test given at a later date.

3) The Level I students in Treatment U were highly successful in handling the relationships among units of area upon the completion of the treatment. An interesting transfer measure would be the repetition of the test involving units of length used in the teach-test procedure.

4) If the treatments or instruments are to be used in further studies, changes should be made according to the remarks found in the conclusion section of this chapter or in the journal in Appendix C. In particular, Treatment N should be strengthened by placing more emphasis on comparing inexactly covered regions.

5) The child's ability to learn about other units of measure needs further investigation. Likewise, the ability of transfer from one attribute to another needs more careful examination. By varying the attributes and obtaining the third level of aptitude the design of this study could be used for such investigations.

6) The teach-test procedure was successful in producing two distinct groups. With the collection of additional data (IQ, teachers' ratings, etc.) the following interpretations of Heimer and Lottes' hypotheses could be tested:

a) This determination of the aptitude levels is a better predictor of success for each of the criterion variables than conventional procedures (IQ, teachers' ratings).

b) This determination of the aptitude levels measured factors not taken into account by conventional procedures for predicting success for each of the criterion variables.

c) This determination of the aptitude levels was a better

predictor of success for each of the criterion variables for Treatment U than for Treatment N.

d) This determination of the aptitude levels was a better predictor of success for each of the criterion variables for Level II than for Level I.

7) Defining aptitude as the ability to learn is not a common approach. Although no significant interactions between aptitude and treatment were found, the investigator feels that this approach is worth pursuing. The treatments and dependent observations need to be refined and the third level of aptitude needs to be obtained, before drawing any conclusions about using this method of determining aptitudes in an ATI study.

8) Performance on Piagetian type tasks related to area and length might profitably be investigated before and after this experiment. Although instruction has not often proved to change performance on Piagetian tasks, the subjects in Treatment U appeared to be coordinating the unit with the number and not centering on only one dimension. This ability may make a difference in the performance on typical area and length Piagetian tasks.

Concluding Remarks

This study was made in response to questions raised by a curriculum development project. Although the hypothesized interaction between aptitudes and treatments was not found, many results were relevant to the development of curriculum materials. Not all questions associated with curriculum development may be answered by research, but many more

questions that are now being answered through research can be and need to be. Thus, there is a need for further research of this type to be made in conjunction with developing curriculum.

REFERENCES

Anderson, G. L. Quantitative thinking as developed under connectionist and field theories of learning. In Esther J. Swenson et al. (Eds.), Learning theory in school situations. Minneapolis: The University of Minnesota Press, 1949.

Becker, J. P. An attempt to design instructional techniques in mathematics to accommodate different patterns of mental ability. Dissertation Abstracts A, 1967, 28, 957-958.

Behr, M. J. Interactions between "Structure-of-Intellect" factors and two methods of presenting concepts of modular arithmetic - a summary paper. Journal for Research in Mathematics Education, 1970, 1, 29-42.

Beilin, H. Perceptual-cognitive conflict in the development of an invariant area concept. Journal of Experimental Child Psychology, 1964, 1, 208-226.

Beilin, H. Feedback and infralogical strategies in invariant area conceptualization. Journal of Experimental Child Psychology, 1966, 3, 267-278.

Beilin, H., & Franklin, I. C. Logical operations in area and length measurement: Age and training effects. Child Development, 1962, 33, 607-618.

Blakers, A. L. Studies in mathematics. Vol. XVII. Mathematical concepts of elementary measurement. Los Angeles: School Mathematics Study Group, 1967.

Bracht, G. H. The relationship of treatment tasks personological variables, and dependent variables to aptitude-treatment interactions. Unpublished doctoral dissertation, University of Colorado, 1969.

Cahen, L. S. Aptitude-treatment interactions: Some conceptual and methodological considerations. Research Bulletin 69-96. Princeton, N. J.: Educational Testing Service, 1969.

Carpenter, T. P. The performance of first grade students on a non-standard set of measurement tasks. Technical Report. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, in press.

Carpenter, T. P. The role of equivalence and order relations in the development and coordination of the concepts of unit size and number of units in selected conservation type measurement problems. Technical Report No. 178. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1971.

Carry, L. R. Interaction of visualization and general reasoning abilities with instructional treatment in algebra. (Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilms, 1968. No. 68-11, 280.

Cronbach, L. J. Evaluation for course improvement. In Richard C. Anderson et al. (Eds.), Current research on instruction. Englewood Cliffs: Prentice-Hall, 1969.

Cronbach, L. J., & Snow, R. E. Final report: Individual differences in learning ability as a function of instructional variables. Stanford: Stanford University, March 1969.

Davidson, E. Mediation and ability in paired-associate learning. Journal of Educational Psychology, 1964, 55, 352-356.

Davis, J. B., Jr. An investigation of the interaction of certain instructional strategies with the structure of basic mental abilities in the learning of some mathematical operations. Dissertation Abstracts A, 1968, 28, 2551-2552.

Ellison, A. The concept of the shifting unit. Arithmetic Teacher, 1972, 19, 171-176.

Gal'perin, P. Ya., & Georgiev, L. S. The formation of elementary mathematical notions. In J. Kilpatrick & I. Wirszup (Eds.) Soviet studies in the psychology of learning and teaching mathematics. Vol. 1, Los Angeles: School Mathematics Study Group, 1969.

Gilbert, L. E. An introduction of length concepts to kindergarten children. Technical Report No. 102. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1969.

Harvey, J. G., Romberg, T. A., & Fletcher, H. J. Analysis of Mathematics Instruction: A discussion and interim report. Paper presented at the First International Congress on Mathematical Education, Lyon, France, 1969.

Hays, W. L. Statistics for Psychologists, New York: Holt, Rinehart & Winston, 1965.

Heimer, R. T. A teach-test procedure for obtaining measures of mathematical aptitude. American Mathematical Monthly, 1966, 73, 883-889.

Heimer, R. T., & Lottes, J. J. Try-out of the teach-test approach in SSTP. (Final Report). NSF Grant GW-1223. University Park: The Pennsylvania State University, 1968.

Huntington, J. R. Linear measurement in the primary grades: A comparison of Piaget's description of the child's spontaneous conceptual development and SMSG sequence of instruction. Journal for Research in Mathematics Education, 1970, 1, 219-232.

Johnson, R. Effects of varying concrete activities on achievement of objectives in perimeter, area, and volume by students of grades four, five, and six. Unpublished doctoral dissertation, University of Colorado, 1970.

Kelly, F. J., Beggs, D. L., & McNeil, K. A. Multiple regression approach. Carbondale: Southern Illinois University Press, 1969.

Kerst, M. W., & Walker, D. F. An analysis of curriculum policy-making. Review of Educational Research, 1971, 41, 479-510.

Klausmeier, H. J., Quilling, M., & Sorenson, J. The development and evaluation of the multiunit elementary school. Technical Report No. 158. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1971.

Lovell, K. Some studies involving spatial ideas. In M. F. Rosskopf, L. P. Steffe, & S. Taback (Eds.), Piagetian cogntive development research and mathematical education. Washington: National Council of Teachers of Mathematics, 1971.

Lovell, K., Healey, D., & Rowland, A. D. Growth of some geometrical concepts. Child Development, 1962, 33, 751-767.

Lubin, A. The interpretation of significant interaction. Educational and Psychological Measurement, 1961, 21, 807-817.

Luchins, A. S., & Luchins, E. H. A structural approach to the teaching of the concept of area in intuitive geometry. Journal of Educational Research, 1947, 40, 528-533.

Lucow, W. An experiment with the Cuisenaire method in grade three. American Educational Research Journal, 1964, 1, 159-167.

Marsh, L. G. Children explore mathematics. (3rd ed.) London: A & C Black, 1969.

Needleman, J. Scalogram analysis of certain area concepts proposed by Piaget. Paper presented at the meeting of the National Council of Mathematics, Anaheim, California, 1971.

Paige, D. D., & Jennings, M. Measurement in the elementary school. Arithmetic Teacher, 1967, 14, 354-357.

Piaget, J., Inhelder, B., & Szeminska, A. The child's conception of geometry. New York: Harper & Row, 1964.

Romberg, T. A., & DeVault, M. V. Mathematics curriculum: Needed research. Journal of Research and Development in Education, 1967, 1, 95-112.

Romberg, T. A., Fletcher, H. J., & Scott, J. A. A measurement approach to elementary mathematics instruction. Working Paper No. 12. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1968.

Romberg, T. A., & Harvey, J. G. Developing mathematical processes: Background and projections. Working Paper No. 14. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1969.

Salomon, G. Heuristics models for the generation of aptitude-treatment interaction hypotheses. Paper presented at the meeting of the American Educational Research Association, New York, 1971.

Scott, J. A. The effect of selected training experiences on performance on a test of conservation of numerousness. Technical Report No. 92. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1969.

Scriven, Michael. The methodology of evaluation. Perspective of curriculum evaluation. AERA Monograph Series on Curriculum Evaluation, 1. Chicago: Rand McNally, 1967.

Stevens, S. S. Measurement, psychophysics, and utility. In C. W. Churchman & P. Ratoosh (Eds.), Measurement: Definitions and theories. New York: John Wiley, 1959.

Suydam, M. N., & Riedesel, C. A. Interpretative study of research and development in elementary school mathematics. Volume 2, Compilation of research reports. Final report, USOE Project No. 8-0586. University Park: The Pennsylvania State University, 1969.

Wagman, G. A study of the child's conception of area measure. Dissertation Abstracts A, 1969, 30, 1350.

Weaver, J. F. Some concerns about the application of Piaget's theory and research in mathematical learning and instruction. Arithmetic Teacher, 1972, 19, 263-270.

Weinstein, M. S. Effects of training on the concepts of water level and horizontality in the classroom. Technical Report 128. Madison: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1970.

Appendices A through F have been ommitted from this publication,
but are available on microfilm from Memorial Library,
University of Wisconsin, Madison, Wisconsin

National Evaluation Committee

Helen Bain
Immediate Past President
National Education Association

Lyle E. Bourne, Jr.
Institute for the Study of Intellectual Behavior
Universiy of Colorado

Jeanne S. Chall
Graduate School of Education
Harvard University

Francis S. Chase
Department of Education
University of Chicago

George E. Dickson
College of Education
University of Toledo

Hugh J. Scott
Superintendent of Public Schools
District of Columbia

H. Craig Sipe
Department of Instruction
State University of New York

G. Wesley Sowards
Dean of Education
Florida International University

Benton J. Underwood
Department of Psychology
Northwestern University

Robert J. Wisner
Mathematics Department
New Mexico State University

Executive Committee

William R. Bush
Director of Program Planning and Management
and Deputy Director, R & D Center

Herbert J. Klausmeier, Committee Chairman
Director, R & D Center

Wayne Otto
Principal Investigator
R & D Center

Robert G. Petzold
Professor of Music
University of Wisconsin

Richard A. Rossmiller
Professor of Educational Administration
University of Wisconsin

James E. Walter
Coordinator of Program Planning
R & D Center

Russell S. Way, ex officio
Program Administrator, Title III ESEA
Wisconsin Department of Public Instruction

Faculty of Principal Investigators

Vernon L. Allen
Professor of Psychology

Frank H. Farley
Associate Professor
Educational Psychology

Marvin J. Fruth
Associate Professor
Educational Administration

John G. Harvey
Associate Professor
Mathematics

Frank H. Hooper
Associate Professor
Child Development

Herbert J. Klausmeier
Center Director
V. A. C. Henmon Professor
Educational Psychology

Stephen J. Knezevich
Professor
Educational Administration

Joel R. Levin
Associate Professor
Educational Psychology

L. Joseph Lins
Professor
Institutional Studies

Wayne Otto
Professor
Curriculum and Instruction

Thomas A. Romberg
Associate Professor
Curriculum and Instruction

Peter A. Schreiber
Assistant Professor
English

Richard L. Venezky
Associate Professor
Computer Science

Alan M. Voelker
Assistant Professor
Curriculum and Instruction

Larry M. Wilder
Assistant Professor
Communication Arts

GPO 826-076